(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 February 2002 (14.02.2002)

PCT

(10) International Publication Number
WO 02/12498 A2

(51) International Patent Classification[7]: C12N 15/12, 5/10, C12Q 1/68, C07K 14/47, 16/18, G06F 17/00

(21) International Application Number: PCT/US01/24664

(22) International Filing Date: 6 August 2001 (06.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/223,535 4 August 2000 (04.08.2000) US

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KLIEM, Stefanie, E.** [DE/DE]; Kiefernweg 37, D-61440 Oberusel (DE). **KOSHY, Beena** [IN/US]; 1298 Hartford Turnpike, Apt. 11B, North Haven, CT 06473 (US). **TANGUAY, Debra, A.** [US/US]; 902 Aspen Glen Drive, Hamden, CT 06518 (US).

(74) Agents: **SHANER, Sandra, L.** et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 02/12498 A2

(54) Title: HAPLOTYPES OF THE ISL1 GENE

(57) Abstract: Novel genetic variants of the ISL1 Transcription Factor, LIM/Homeodomain, (islet-1) (ISL1) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the ISL1 gene. Compositions and methods for haplotyping and/or genotyping the ISL1 gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.

# HAPLOTYPES OF THE ISL1 GENE

## RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/223,535 filed August 4, 2000.

## FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

## BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of such drugs in clinical

trials or their early withdrawal from the market even though they could be highly beneficial for other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* **15**:1249-52; Kleyn PW et al. 1998 *Science* **281**: 1820-21; Kola I 1999 *Curr Opin Biotech* **10**:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* **66**:445-7; Marshall, E 1999 *Science* **284**:406-7; Judson R et al. 2000 *Pharmacogenomics* **1**:1-12; Roses AD 2000 *Nature* **405**:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* **19**:216-7; Wang DG et al 1998 *Science* **280**:1077-82; Chakravarti A 1999 *Nat Genet* **21**:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* **4**:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* **5**:538-43; Davidson S 2000 *Nature Biotech* **18**:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* **161**: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* **320**:987-90; Dahl BS 1997 *Acta Psychiatr Scand* **96** (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* **63**:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* **97**:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of disorders related to defects in motor neuron and diabetes is the ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) gene or its encoded product. ISL1 is a member of the LIM/homeodomain family of transcription factors. Initial experiments localized ISL1 to the islet cells of the pancreas, suggesting a possible role for this protein in the regulation of insulin gene expression and/or islet cell development. More recent studies have determined that ISL1 is also expressed in the central and peripheral nervous system (Wang and Drucker

*Endocrinology* 1994;134(3):1416-22).

Gene deletion studies in mice have revealed that both LIM homeodomain and LMO proteins are essential for embryonic viability, as well as for the development of motor neurons and interneurons, the pancreas, the pituitary, and erythrocytes (Jurata et al. *J Biol Chem* 1998, 273:3152-7). Dorsal pancreatic mesenchyme does not form in ISL1-mutant embryos and there is an associated failure of exocrine cell differentiation in the dorsal but not the ventral pancreas. There is also a complete loss of differentiated islet cells. These results indicate that ISL1 is necessary for the development of the dorsal exocrine pancreas, and also that ISL1 function in pancreatic endodermal cells is required for the generation of all endocrine islet cells.

The combinatorial expression of the LIM homeodomain proteins ISL1, ISL2, LHX1, and LHX3 in subsets of developing motor neurons correlates with the future organization of these neurons into motor columns with distinct innervation targets, implying a functional role for LIM homeodomain protein combinations in the specification of neuronal identity. In studies analyzing cell differentiation in the neural tube of embryos in which ISL1 expression has been eliminated by gene targeting, Pfaff et al. (*Cell* 1996 84:309-20) showed that motor neurons are not generated without ISL1, although many other aspects of cell differentiation in the neural tube occur normally. These results show that ISL1 is required for the generation of motor neurons and suggest that motor neuron generation is required for the subsequent differentiation of certain interneurons.

The ISL1 transcription factor, LIM/homeodomain, (islet-1) gene is located on chromosome 5q and contains 6 exons that encode a 346 amino acid protein. A reference sequence for the ISL1 gene is shown in the contiguous lines of Figure 1 (Genaissance Reference No. 401083; SEQ ID NO: 1). Reference sequences for the coding sequence (GenBank Accession No. NM_002202.1) and protein are shown in Figures 2 (SEQ ID NO: 2) and 3 (SEQ ID NO: 3), respectively.

Because of the potential for variation in the ISL1 gene to affect the expression and function of the encoded protein, it would be useful to know whether polymorphisms exist in the ISL1 gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of ISL1 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

## SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 14 novel polymorphic sites in the ISL1 gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 3542 (PS1), 3732 (PS2), 3789 (PS3), 3794 (PS4), 3987 (PS5), 4078 (PS6), 4084 (PS7), 7837 (PS8), 7885 (PS9), 8183 (PS10), 10193 (PS11), 11502 (PS12), 13835 (PS13) and 14055 (PS14). The polymorphisms at these sites are adenine or guanine at PS1, adenine or guanine at PS2, guanine or cytosine at PS3, guanine or adenine at PS4, adenine or guanine at PS5, cytosine or thymine at PS6, guanine or adenine at PS7, cytosine or guanine at PS8, cytosine or guanine at PS9, thymine or cytosine at PS10, cytosine or

thymine at PS11, guanine or adenine at PS12, adenine or guanine at PS13 and guanine or adenine at PS14. In addition, the inventors have determined the identity of the alleles at these sites in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS14 in the ISL1 gene, which are shown below in Tables 5 and 4, respectively. Each of these ISL1 haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the ISL1 gene that exists in the human population. The frequency with which each haplotype and haplotype pair occurs within the total reference population and within each of the four major population groups included in the reference population was also determined.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the ISL1 gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13 and PS14 in both copies of the ISL1 gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel ISL1 polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel ISL1 polymorphic sites. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 5 below or has one of the haplotype pairs in Table 4 below.

The invention also provides a method for haplotyping the ISL1 gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the ISL1 gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13 and PS14. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's ISL1 gene is defined by one of the ISL1 haplotypes shown in Table 5, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's ISL1 gene are defined by one of the ISL1 haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof. The method for establishing the ISL1 haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with ISL1 activity, e.g., disorders related to defects in motor neuron and diabetes.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate ISL1 as a candidate target for treating a specific condition or disease predicted to be associated with ISL1 activity. Determining for a particular population the frequency of one or more of the individual ISL1 haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue ISL1 as a target for treating the specific disease of interest. In particular, if variable ISL1

activity is associated with the disease, then one or more ISL1 haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed ISL1 haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable ISL1 activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any ISL1 haplotype or haplotype pair, apply the information derived from detecting ISL1 haplotypes in an individual to decide whether modulating ISL1 activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting ISL1 to treat a specific condition or disease predicted to be associated with ISL1 activity. For example, detecting which of the ISL1 haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the most frequent ISL1 isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular ISL1 haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

The method for haplotyping the ISL1 gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with ISL1 activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the ISL1 haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute ISL1 haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of a ISL1 haplotype or haplotype pair that had a previously unknown association with response to the drug being studied in the trial. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any ISL1 haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and a ISL1 genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the ISL1 genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the ISL1 genotype or haplotype in a reference population. A higher frequency of the ISL1 genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the ISL1 genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the ISL1 haplotype is selected from the haplotypes shown in Table 5, or a sub-haplotype thereof. Such

methods have applicability in developing diagnostic tests and therapeutic treatments for disorders related to defects in motor neuron and diabetes.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the ISL1 gene or a fragment thereof. The reference sequence comprises the contiguous sequences shown in Figure 1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of guanine at PS1, guanine at PS2, cytosine at PS3, adenine at PS4, guanine at PS5, thymine at PS6, adenine at PS7, guanine at PS8, guanine at PS9, cytosine at PS10, thymine at PS11, adenine at PS12, guanine at PS13 and adenine at PS14.

A particularly preferred polymorphic variant is an isogene of the ISL1 gene. A ISL1 isogene of the invention comprises adenine or guanine at PS1, adenine or guanine at PS2, guanine or cytosine at PS3, guanine or adenine at PS4, adenine or guanine at PS5, cytosine or thymine at PS6, guanine or adenine at PS7, cytosine or guanine at PS8, cytosine or guanine at PS9, thymine or cytosine at PS10, cytosine or thymine at PS11, guanine or adenine at PS12, adenine or guanine at PS13 and guanine or adenine at PS14. The invention also provides a collection of ISL1 isogenes, referred to herein as a ISL1 genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a ISL1 cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of guanine at a position corresponding to nucleotide 243 and thymine at a position corresponding to nucleotide 655. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a ISL1 isogene defined by haplotypes 2c and 7c.

Polynucleotides complementary to these ISL1 genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the ISL1 gene will be useful in studying the expression and function of ISL1, and in expressing ISL1 protein for use in screening for candidate drugs to treat diseases related to ISL1 activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express ISL1 for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the ISL1 protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of methionine at a position corresponding to amino acid position 81 and serine at a position corresponding to amino acid position 219. A polymorphic variant of ISL1 is useful in studying the effect of the variation on the biological activity of ISL1 as well as on the binding affinity of candidate drugs targeting ISL1 for the treatment of disorders related to defects in motor

neuron and diabetes.

The present invention also provides antibodies that recognize and bind to the above polymorphic ISL1 protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one of the ISL1 polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the ISL1 isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against ISL1 protein, and for testing the efficacy of therapeutic agents and compounds for disorders related to defects in motor neuron and diabetes in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the ISL1 gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the ISL1 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing ISL1 haplotypes organized according to their evolutionary relationships.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the ISL1 gene (Genaissance Reference No. 401083; contiguous lines), with the start and stop positions of each region of coding sequence indicated with a bracket ([ or ]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:1 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25). SEQ ID NO:74 is a modified version of SEQ ID NO:1 that shows the context sequence of each polymorphic site, PS1-PS14, in a uniform format to facilitate electronic searching. For each polymorphic site, SEQ ID NO:74 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30$^{th}$ position, followed by 60 bases of unspecified sequence to represent that each PS is separated by genomic sequence whose composition is defined elsewhere herein.

Figure 2 illustrates a reference sequence for the ISL1 coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the ISL1 protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the ISL1 gene. As described in more detail below, the inventors herein discovered 15 isogenes of the ISL1 gene by characterizing the ISL1 gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (21 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference population were organized into population subgroups by their self-identified ethnogeographic origin as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

| Population Group | Population Subgroup | No. of Individuals |
|---|---|---|
| African descent | | 20 |
| | Sierra Leone | 1 |
| Asian | | 20 |
| | Burma | 1 |
| | China | 3 |
| | Japan | 6 |
| | Korea | 1 |
| | Philippines | 5 |
| | Vietnam | 4 |
| Caucasian | | 21 |
| | British Isles | 3 |
| | British Isles/Central | 4 |
| | British Isles/Eastern | 1 |
| | Central/Eastern | 1 |
| | Eastern | 3 |
| | Central/Mediterranean | 1 |
| | Mediterranean | 2 |
| | Scandinavian | 2 |
| Hispanic/Latino | | 18 |
| | Caribbean | 8 |
| | Caribbean (Spanish Descent) | 2 |
| | Central American (Spanish Descent) | 1 |
| | Mexican American | 4 |
| | South American (Spanish Descent) | 3 |

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The ISL1 isogenes present in the human reference population are defined by haplotypes for 14 polymorphic sites in the ISL1 gene, all of which are believed to be novel. The novel ISL1 polymorphic sites identified by the inventors are referred to as PS1-PS14 to designate the order in which they are located in the gene (see Table 3 below). Using the genotypes identified in the Index Repository for

PS1-PS14 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the ISL1 gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the ISL1 gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether ISL1 is a suitable target for drugs to treat disorders related to defects in motor neuron and diabetes, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

**Allele** - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

**Candidate Gene** – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

**Gene** - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

**Genotype** – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

**Full-genotype** – The unphased 5′ to 3′ sequence of nucleotide pairs found at all polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

**Sub-genotype** – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

**Genotyping** – A process for determining a genotype of an individual.

**Haplotype** – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

**Full-haplotype** – The 5′ to 3′ sequence of nucleotides found at all polymorphic sites examined herein in a locus on a single chromosome from a single individual.

**Sub-haplotype** – The 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a single chromosome from a single individual.

**Haplotype pair** – The two haplotypes found for a locus in a single individual.

**Haplotyping** – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

**Haplotype data** - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations

between one or more haplotypes and a trait.

**Isoform** – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

**Isogene** – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

**Isolated** – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

**Locus** - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature, wherein the physical feature may be a polymorphic site.

**Naturally-occurring** – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

**Nucleotide pair** – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

**Phased** – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

**Polymorphic site (PS)** – A position within a locus at which at least two alternative sequences are found in a population.

**Polymorphic variant** – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

**Polymorphism** – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

**Polymorphism data** – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

**Polymorphism Database** – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

**Polynucleotide** – A nucleic acid molecule comprised of single-stranded RNA or DNA or

comprised of complementary, double-stranded DNA.

**Population Group** – A group of individuals sharing a common ethnogeographic origin.

**Reference Population** – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

**Single Nucleotide Polymorphism (SNP)** – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

**Subject** – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

**Treatment** - A stimulus administered internally or externally to a subject.

**Unphased** – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the ISL1 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel ISL1 polymorphisms and haplotypes identified herein.

The compositions comprise at least one ISL1 genotyping oligonucleotide. In one embodiment, a ISL1 genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a ISL1 polynucleotide, i.e., a ISL1 isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-ISL1 polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the ISL1 gene using the polymorphism information provided herein in conjunction with the known sequence information for the ISL1 gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central

position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention.

ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent that the ASO contains one of the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting ISL1 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

CAGCCCGRGCTTCGG (SEQ ID NO:4) and its complement,
CTGCGTCRGACCAAT (SEQ ID NO:5) and its complement,
TAAGGAASAGAGGTG (SEQ ID NO:6) and its complement,
AAGAGAGRTGCCCGA (SEQ ID NO:7) and its complement,
TTCACCARCTGTACA (SEQ ID NO:8) and its complement,
TATTTTAYCTTGTGG (SEQ ID NO:9) and its complement,
ACCTTGTRGGGCTCG (SEQ ID NO:10) and its complement,
CGCTCCCSCCTCCCC (SEQ ID NO:11) and its complement,
GCAGCATSGGCTTCA (SEQ ID NO:12) and its complement,
CAACAACYATGGTAG (SEQ ID NO:13) and its complement,
CCTCAGTYCCCGTGT (SEQ ID NO:14) and its complement,
GTACGGCRGATTAAC (SEQ ID NO:15) and its complement,
TTTATTTRTTTCTCA (SEQ ID NO:16) and its complement, and
ATTTAACRACCCAGT (SEQ ID NO:17) and its complement.

A preferred ASO primer for detecting ISL1 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

GCCCGGCAGCCCGRG (SEQ ID NO:18); TCTCCCCCGAAGCYC (SEQ ID NO:19);
CGCGCGCTGCGTCRG (SEQ ID NO:20); ATCGCCATTGGTCYG (SEQ ID NO:21);
AAGAGATAAGGAASA (SEQ ID NO:22); CTCGGGCACCTCTST (SEQ ID NO:23);
ATAAGGAAGAGAGRT (SEQ ID NO:24); CGCGGCTCGGGCAYC (SEQ ID NO:25);
GGGCTGTTCACCARC (SEQ ID NO:26); GGTGGTTGTACAGYT (SEQ ID NO:27);
AGAGGCTATTTTAYC (SEQ ID NO:28); CGAGCCCCACAAGRT (SEQ ID NO:29);
TATTTTACCTTGTRG (SEQ ID NO:30); GCACACCGAGCCCYA (SEQ ID NO:31);
CGCCTCCGCTCCCSC (SEQ ID NO:32); TGTGCGGGGGAGGSG (SEQ ID NO:33);
CCAAGTGCAGCATSG (SEQ ID NO:34); TCTTGCTGAAGCCSA (SEQ ID NO:35);
CGGCCACAACAACYA (SEQ ID NO:36); CTGTAGCTACCATRG (SEQ ID NO:37);
GACGGGCCTCAGTYC (SEQ ID NO:38); CGGATCACACGGGRA (SEQ ID NO:39);
CCGTAGGTACGGCRG (SEQ ID NO:40); GACTCAGTTAATCYG (SEQ ID NO:41);
GTGTCCTTTATTTRT (SEQ ID NO:42); GAAGGTTGAGAAAYA (SEQ ID NO:43);
AAAAGTATTTAACRA (SEQ ID NO:44); and TCATTGACTGGGTYG (SEQ ID NO:45).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting ISL1 gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

CGGCAGCCCG (SEQ ID NO:46); CCCCCGAAGC (SEQ ID NO:47);
GCGCTGCGTC (SEQ ID NO:48); GCCATTGGTC (SEQ ID NO:49);
AGATAAGGAA (SEQ ID NO:50); GGGCACCTCT (SEQ ID NO:51);
AGGAAGAGAG (SEQ ID NO:52); GGCTCGGGCA (SEQ ID NO:53);
CTGTTCACCA (SEQ ID NO:54); GGTTGTACAG (SEQ ID NO:55);
GGCTATTTTA (SEQ ID NO:56); GCCCCACAAG (SEQ ID NO:57);
TTTACCTTGT (SEQ ID NO:58); CACCGAGCCC (SEQ ID NO:59);
CTCCGCTCCC (SEQ ID NO:60); GCGGGGGAGG (SEQ ID NO:61);
AGTGCAGCAT (SEQ ID NO:62); TGCTGAAGCC (SEQ ID NO:63);
CCACAACAAC (SEQ ID NO:64); TAGCTACCAT (SEQ ID NO:65);
GGGCCTCAGT (SEQ ID NO:66); ATCACACGGG (SEQ ID NO:67);
TAGGTACGGC (SEQ ID NO:68); TCAGTTAATC (SEQ ID NO:69);
TCCTTTATTT (SEQ ID NO:70); GGTTGAGAAA (SEQ ID NO:71);
AGTATTTAAC (SEQ ID NO:72); and TTGACTGGGT (SEQ ID NO:73).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

ISL1 genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized ISL1 genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer

extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the ISL1 gene in an individual. As used herein, the terms "ISL1 genotype" and "ISL1 haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the ISL1 gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid sample comprising the two copies of the ISL1 gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13 and PS14 in the two copies to assign a ISL1 genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-PS14.

Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the ISL1 gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ untranslated regions. If a ISL1 gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the ISL1 gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13 and PS14 in that copy to assign a ISL1 haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the ISL1 gene or fragment such as one of the methods described above for preparing ISL1 isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two ISL1 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional ISL1 clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the ISL1 gene in an individual. In a particularly preferred embodiment, the nucleotide at each of PS1-PS14 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the ISL1 haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual's ISL1 gene, the phased sequence of nucleotides present at each of PS1-PS14. The present invention also contemplates that typically only a subset of PS1-PS14 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 5. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, a ISL1 haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13 and PS14 in each copy of the ISL1 gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS14 in each copy of the ISL1 gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the ISL1 gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not

guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the ISL1 gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize

nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's ISL1 haplotype pair is predicted from its ISL1 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a ISL1 genotype for the individual at two or more ISL1 polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing ISL1 haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the ISL1 haplotype pairs shown in Table 4.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For

example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by 2n=log(1-q)/log(1-p) where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair $H_1 / H_2$ is equal to $p_{H\text{-}W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H\text{-}W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting a ISL1 haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology

(U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*). A preferred process for predicting ISL1 haplotype pairs from ISL1 genotypes is described in U.S. Provisional Application Serial No. 60/198,340 and the corresponding International Application, PCT/US01/12831.

The invention also provides a method for determining the frequency of a ISL1 genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel ISL1 polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be a reference population, a family population, a same sex population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for ISL1 genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and a ISL1 genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular ISL1 genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that ISL1 genotype, haplotype or haplotype pair. Preferably, the ISL1 genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting ISL1 or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other

disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a ISL1 genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the ISL1 gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and ISL1 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their ISL1 genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the ISL1 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in PCT Application

Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between ISL1 haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra*, Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in PCT Application Serial No. PCT/US00/17540.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the ISL1 gene. As described in PCT Application Serial No. PCT/US00/17540, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of ISL1 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the ISL1 gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the ISL1 gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying ISL1 genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the ISL1 gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant ISL1 gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13 and PS14. Similarly, the nucleotide sequence of a variant fragment of the ISL1 gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic

sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the ISL1 gene, which is defined by haplotype 5, (or other reported ISL1 sequences) or to portions of the reference sequence (or other reported ISL1 sequences), except for genotyping oligonucleotides as described above.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of guanine at PS1, guanine at PS2, cytosine at PS3, adenine at PS4, guanine at PS5, thymine at PS6, adenine at PS7, guanine at PS8, guanine at PS9, cytosine at PS10, thymine at PS11, adenine at PS12, guanine at PS13 and adenine at PS14. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the ISL1 gene which is defined by any one of haplotypes 1- 4 and 6 - 15 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the ISL1 gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

ISL1 isogenes may be isolated using any method that allows separation of the two "copies" of the ISL1 gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides ISL1 genome anthologies, which are collections of ISL1 isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. A ISL1 genome anthology may comprise individual ISL1 isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the ISL1 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred ISL1 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant

expression vector capable of being propagated and expressing the encoded ISL1 protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant ISL1 sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the ISL1 gene will produce ISL1 mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a ISL1 cDNA comprising a nucleotide sequence which is a polymorphic variant of the ISL1 reference coding sequence shown in Figure 2. Thus, the invention also provides ISL1 mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of guanine at a position corresponding to nucleotide 243 and thymine at a position corresponding to nucleotide 655. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a ISL1 isogene defined by haplotypes 2c and 7c. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to

previously identified and characterized ISL1 cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of a ISL1 gene fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the ISL1 polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the ISL1 gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the ISL1 genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular ISL1 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the ISL1 isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular ISL1 isogene. Expression of a ISL1 isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions –10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of ISL1 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of ISL1 mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to

increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference ISLI amino acid sequence shown in Figure 3. The location of a variant amino acid in a ISL1 polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig. 3). A ISL1 protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 except for having one or more variant amino acids selected from the group consisting of methionine at a position corresponding to amino acid position 81 and serine at a position corresponding to amino acid position 219. The invention specifically excludes amino acid sequences identical to those previously identified for ISL1, including SEQ ID NO:3, and previously described fragments thereof. ISL1 protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, a ISL1 protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes shown in Table 5.

Table 2. Novel Polymorphic Variants of ISL1

| Polymorphic Variant Number | Amino Acid Position and Identities 81 | 219 |
|---|---|---|
| 1 | I | S |
| 2 | M | P |
| 3 | M | S |

The invention also includes ISL1 peptide variants, which are any fragments of a ISL1 protein variant that contain one or more of the amino acid variations shown in Table 2 A ISL1 peptide variant is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such ISL1 peptide variants may be useful as antigens to generate antibodies specific for one of the above ISL1 isoforms. In addition, the ISL1 peptide variants may be useful in drug screening assays.

A ISL1 variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant ISL1 genomic and cDNA sequences as described above. Alternatively, the ISL1 protein variant may be isolated from a biological sample of an individual having a ISL1 isogene which encodes the variant protein. Where the sample contains two different ISL1 isoforms (i.e., the individual has different ISL1 isogenes), a particular ISL1 isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular ISL1 isoform but does not bind to the other ISL1 isoform.

The expressed or isolated ISL1 protein may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the

isoform of the ISL1 protein as discussed further below. ISL1 variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant ISL1 gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric ISL1 protein. The non-ISL1 portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the ISL1 and non-ISL1 portions so that the ISL1 protein may be cleaved and purified away from the non-ISL1 portion.

An additional embodiment of the invention relates to using a novel ISL1 protein isoform in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known ISL1 protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The ISL1 protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to a ISL1 variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the ISL1 protein(s) of interest and then washed. Bound ISL1 protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel ISL1 protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the ISL1 protein.

In yet another embodiment, when a particular ISL1 haplotype or group of ISL1 haplotypes encodes a ISL1 protein variant with an amino acid sequence distinct from that of ISL1 protein isoforms encoded by other ISL1 haplotypes, then detection of that particular ISL1 haplotype or group of ISL1 haplotypes may be accomplished by detecting expression of the encoded ISL1 protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel ISL1 variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The ISL1 protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the ISL1 protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel protein isoforms described herein is administered to an individual to neutralize activity of the ISL1 isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the ISL1 protein variant from solution as well as react with ISL1 protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect ISL1 protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel ISL1 protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the ISL1 protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: Laboratory Techniques in Biochemistry and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989,

Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci.USA 86;10029).

Effect(s) of the polymorphisms identified herein on expression of ISL1 may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the ISL1 gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into ISL1 protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired ISL1 isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the ISL1 isogene is introduced into a cell in such a way that it recombines with the endogenous ISL1 gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired ISL1 gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the ISL1 isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the ISL1 isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant ISL1 gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the ISL1 isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human ISL1 isogene and producing human ISL1

protein can be used as biological models for studying diseases related to abnormal ISL1 expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel ISL1 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel ISL1 isogenes; an antisense oligonucleotide directed against one of the novel ISL1 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel ISL1 isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel ISL1 isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the ISL1 gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on

ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The ISL1 polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

## EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

## EXAMPLE 1

This example illustrates examination of various regions of the ISL1 gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the ISL1 gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1.

PCR Primer Pairs

| Fragment No. | Forward Primer | Reverse Primer | PCR Product |
|---|---|---|---|
| Fragment 1 | 3264-3283 | complement of 3924-3904 | 661 nt |
| Fragment 2 | 3480-3502 | complement of 4167-4145 | 688 nt |
| Fragment 3 | 3582-3602 | complement of 4132-4111 | 551 nt |
| Fragment 4 | 3583-3602 | complement of 4277-4256 | 695 nt |
| Fragment 5 | 7653-7675 | complement of 8122-8101 | 470 nt |
| Fragment 6 | 7848-7870 | complement of 8333-8312 | 486 nt |
| Fragment 7 | 9853-9874 | complement of 10596-10574 | 744 nt |
| Fragment 8 | 11423-11445 | complement of 11985-11964 | 563 nt |
| Fragment 9 | 13651-13673 | complement of 14430-14408 | 780 nt |

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

| | |
|---|---|
| Reaction volume | = 10 µl |
| 10 x Advantage 2 Polymerase reaction buffer (Clontech) | = 1 µl |
| 100 ng of human genomic DNA | = 1 µl |
| 10 mM dNTP | = 0.4 µl |
| Advantage 2 Polymerase enzyme mix (Clontech) | = 0.2 µl |
| Forward Primer (10 µM) | = 0.4 µl |
| Reverse Primer (10 µM) | = 0.4 µl |
| Water | = 6.6µl |

Amplification profile:
97°C - 2 min. 1 cycle

97°C - 15 sec.
70°C - 45 sec.
72°C - 45 sec. } 10 cycles

97°C - 15 sec.
64°C - 45 sec.
72°C - 45 sec. } 35 cycles

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using the primer sets described previously or those represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

| Fragment No. | Forward Primer | Reverse Primer |
|---|---|---|
| Fragment 1 | 3324-3344 | complement of 3863-3844 |
| Fragment 2 | 3583-3602 | complement of 4122-4103 |
| Fragment 3 | 3637-3655 | complement of 4093-4074 |
| Fragment 4 | 3741-3760 | complement of 4209-4190 |
| Fragment 5 | 7727-7747 | complement of 8091-8072 |
| Fragment 6 | 7882-7901 | complement of 8264-8245 |
| Fragment 7 | 10053-10072 | complement of 10545-10526 |
| Fragment 8 | 11488-11507 | complement of 11930-11911 |
| Fragment 9 | 13767-13787 | complement of 14234-14215 |

Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the ISL1 gene are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the ISL1 Gene

| Polymorphic Site Number | PolyId[a] | Nucleotide Position | Reference Allele | Variant Allele | CDS Variant Position | AA Variant |
|---|---|---|---|---|---|---|
| PS1 | 401111 | 3542 | A | G | | |
| PS2 | 401115 | 3732 | A | G | | |
| PS3 | 401117 | 3789 | G | C | | |
| PS4 | 401119 | 3794 | G | A | | |
| PS5 | 401123 | 3987 | A | G | | |
| PS6 | 401125 | 4078 | C | T | | |
| PS7 | 401127 | 4084 | G | A | | |
| PS8 | 401131 | 7837 | C | G | | |
| PS9 | 401133 | 7885 | C | G | 243 | I81M |
| PS10 | 401139 | 8183 | T | C | | |
| PS11 | 401143 | 10193 | C | T | 655 | P219S |
| PS12 | 401153 | 11502 | G | A | | |
| PS13 | 401157 | 13835 | A | G | | |
| PS14 | 401159 | 14055 | G | A | | |

[a]PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.

EXAMPLE 2

This example illustrates analysis of the ISL1 polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 4(Part1). Genotypes and Haplotype Pairs Observed for ISL1 Gene

| Genotype Number | Polymorphic Sites PS1 | PS2 | PS3 | PS4 | PS5 | PS6 | PS7 | PS8 | PS9 | PS10 | HAP | Pair |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A | A | G | G | A | C | G | C | C | C | 3 | 3 |
| 2 | A | A | G | G | A | C | G | C | C | T | 5 | 5 |
| 3 | A/G | A | G | G | A | C/T | G | C | C | C | 3 | 13 |
| 4 | A | A | G | G | A | C/T | G | C | C | T/C | 5 | 8 |
| 5 | A/G | A | G/C | G | A | C | G | C | C | T/C | 5 | 9 |
| 6 | A/G | A | G | G | A/G | C | G | C | C | T/C | 5 | 14 |
| 7 | A | A | G | G | A | C | G | C/G | C | T | 5 | 7 |
| 8 | A | A | G | G | A | C | G | C | C | T | 5 | 6 |
| 9 | A | A | G | G/A | A | C | G | C | C | T | 5 | 1 |
| 10 | A/G | A/G | G | G | A | C | G | C | C | C | 3 | 15 |
| 11 | A | A | G | G/A | A | C | G | C | C/G | T/C | 5 | 2 |
| 12 | G/A | A | G | G | A | C | G | C | C | C/T | 12 | 4 |
| 13 | A/G | A | G | G | A/G | C | G | C | C | C | 3 | 14 |
| 14 | A/G | A | G | G | A | C | G/A | C | C | C | 3 | 10 |
| 15 | A/G | A | G | G | A | C | G | C | C | C | 3 | 12 |
| 16 | A/G | A | G | G | A | C | G | C | C | T/C | 5 | 12 |
| 17 | A | A | G | G | A | C | G | C | C | T/C | 5 | 3 |
| 18 | A/G | A | G | G | A | C | G | C | C | C | 3 | 11 |

Table 4(Part2). Genotypes and Haplotype Pairs Observed for ISL1 Gene

| Genotype Number | Polymorphic Sites PS11 | PS12 | PS13 | PS14 | HAP | Pair |
|---|---|---|---|---|---|---|
| 1 | C | G | A | G | 3 | 3 |
| 2 | C | G | A | G | 5 | 5 |
| 3 | C | G | A | G | 3 | 13 |
| 4 | C | G | A | G | 5 | 8 |
| 5 | C | G | A | G | 5 | 9 |
| 6 | C | G | A | G | 5 | 14 |
| 7 | C/T | G | A | G | 5 | 7 |
| 8 | C | G | A/G | G | 5 | 6 |
| 9 | C | G | A | G | 5 | 1 |
| 10 | C | G | A | G | 3 | 15 |
| 11 | C | G | A | G | 5 | 2 |
| 12 | C | G | A | G/A | 12 | 4 |
| 13 | C | G | A | G | 3 | 14 |
| 14 | C | G | A | G | 3 | 10 |
| 15 | C | G | A | G | 3 | 12 |
| 16 | C | G | A | G | 5 | 12 |
| 17 | C | G | A | G | 5 | 3 |
| 18 | C | G/A | A | G | 3 | 11 |

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample, as described in U.S. Provisional Application Serial No. 60/198,340 entitled "A Method and System for Determining Haplotypes from a Collection of Polymorphisms" and the corresponding International Application, PCT/US01/12831. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals. In our

analysis, the list of haplotypes was augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family).

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 15 human ISL1 haplotypes shown in Table 5 below.

An ISL1 isogene defined by a full-haplotype shown in Table 5 below comprises the regions of the SEQ ID NOS indicated in Table 5, with their corresponding set of polymorphic locations and identities, which are also set forth in Table 5.

Table 5. Haplotypes Identified in the ISL1 Gene

| Haplotype Number[a] | | | | | | | | | | PS No.[b] | PS Pos.[c] | SEQ ID No.[d] | Region Examined[e] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | | | |
| A | A | A | A | A | A | A | A | G | G | 1 | 3542/30 | 1/74 | 3264-4277 |
| A | A | A | A | A | A | A | A | A | A | 2 | 3732/150 | 1/74 | 3264-4277 |
| G | G | G | G | G | G | G | G | C | G | 3 | 3789/270 | 1/74 | 3264-4277 |
| A | A | G | G | G | G | G | G | G | G | 4 | 3794/390 | 1/74 | 3264-4277 |
| A | A | A | A | A | A | A | A | A | A | 5 | 3987/510 | 1/74 | 3264-4277 |
| C | C | C | C | C | C | C | T | C | C | 6 | 4078/630 | 1/74 | 3264-4277 |
| G | G | G | G | G | G | G | G | G | A | 7 | 4084/750 | 1/74 | 3264-4277 |
| C | C | C | C | C | C | G | C | C | C | 8 | 7837/870 | 1/74 | 7653-8333 |
| C | G | C | C | C | C | C | C | C | C | 9 | 7885/990 | 1/74 | 7653-8333 |
| T | C | C | T | T | T | T | C | C | C | 10 | 8183/1110 | 1/74 | 7653-8333 |
| C | C | C | C | C | C | T | C | C | C | 11 | 10193/1230 | 1/74 | 9853-10596 |
| G | G | G | G | G | G | G | G | G | G | 12 | 11502/1350 | 1/74 | 11423-11985 |
| A | A | A | A | A | G | A | A | A | A | 13 | 13835/1470 | 1/74 | 13651-14430 |
| G | G | G | A | G | G | G | G | G | G | 14 | 14055/1590 | 1/74 | 13651-14430 |

| Haplotype Number[a] | | | | | PS No.[b] | PS Pos.[c] | SEQ ID No.[d] | Region Examined[e] |
|---|---|---|---|---|---|---|---|---|
| 11 | 12 | 13 | 14 | 15 | | | | |
| G | G | G | G | G | 1 | 3542/30 | 1/74 | 3264-4277 |
| A | A | A | A | G | 2 | 3732/150 | 1/74 | 3264-4277 |
| G | G | G | G | G | 3 | 3789/270 | 1/74 | 3264-4277 |
| G | G | G | G | G | 4 | 3794/390 | 1/74 | 3264-4277 |
| A | A | A | G | A | 5 | 3987/510 | 1/74 | 3264-4277 |
| C | C | T | C | C | 6 | 4078/630 | 1/74 | 3264-4277 |
| G | G | G | G | G | 7 | 4084/750 | 1/74 | 3264-4277 |
| C | C | C | C | C | 8 | 7837/870 | 1/74 | 7653-8333 |
| C | C | C | C | C | 9 | 7885/990 | 1/74 | 7653-8333 |
| C | C | C | C | C | 10 | 8183/1110 | 1/74 | 7653-8333 |
| C | C | C | C | C | 11 | 10193/1230 | 1/74 | 9853-10596 |
| A | G | G | G | G | 12 | 11502/1350 | 1/74 | 11423-11985 |
| A | A | A | A | A | 13 | 13835/1470 | 1/74 | 13651-14430 |
| G | G | G | G | G | 14 | 14055/1590 | 1/74 | 13651-14430 |

[a]Alleles for ISL1 haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS within the indicated SEQ ID NO, with the 1st position number referring to the first SEQ ID NO and the 2nd position number referring to the 2nd SEQ ID NO;
[d]1st SEQ ID NO refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol; 2nd SEQ ID NO is a modified version of the 1st SEQ ID NO that comprises the context sequence of each polymorphic site, PS1-PS14, to facilitate electronic searching of the haplotypes;
[e]Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

SEQ ID NO:1 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol. SEQ ID NO:74 is a modified version of SEQ ID NO:1 that shows the context sequence of each of PS1-PS14 in a uniform format to facilitate electronic searching of the ISL1 haplotypes. For each polymorphic site, SEQ ID NO:74contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each polymorphic site is separated by genomic sequence whose composition is defined elsewhere herein.

Table 6 below shows the percent of chromosomes characterized by a given ISL1 haplotype for all unrelated individuals in the Index Repository for which haplotype data was obtained. The percent of these unrelated individuals who have a given ISL1 haplotype pair is shown in Table 7. In Tables 6 and 7, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according to their self-identified ethnogeographic origin. Abbreviations used in Tables 6 and 7 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and AM = Native American.

Table 6. Frequency of Observed ISL1 Haplotypes In Unrelated Individuals

| HAP No. | HAP ID | Total | CA | AF | AS | HL | AM |
|---|---|---|---|---|---|---|---|
| 1 | 403730 | 0.61 | 2.38 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 403728 | 0.61 | 2.38 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 403718 | 32.32 | 21.43 | 47.5 | 20.0 | 41.67 | 33.33 |
| 4 | 403724 | 0.61 | 2.38 | 0.0 | 0.0 | 0.0 | 0.0 |
| 5 | 403717 | 51.83 | 61.9 | 25.0 | 72.5 | 44.44 | 66.67 |
| 6 | 403731 | 0.61 | 0.0 | 0.0 | 0.0 | 2.78 | 0.0 |
| 7 | 403727 | 0.61 | 0.0 | 0.0 | 2.5 | 0.0 | 0.0 |
| 8 | 403722 | 0.61 | 0.0 | 0.0 | 0.0 | 2.78 | 0.0 |
| 9 | 403725 | 0.61 | 0.0 | 0.0 | 2.5 | 0.0 | 0.0 |
| 10 | 403726 | 0.61 | 0.0 | 2.5 | 0.0 | 0.0 | 0.0 |
| 11 | 403723 | 0.61 | 0.0 | 2.5 | 0.0 | 0.0 | 0.0 |
| 12 | 403719 | 7.93 | 4.76 | 20.0 | 2.5 | 5.56 | 0.0 |
| 13 | 403732 | 0.61 | 0.0 | 0.0 | 0.0 | 2.78 | 0.0 |
| 14 | 403720 | 1.22 | 2.38 | 2.5 | 0.0 | 0.0 | 0.0 |
| 15 | 403729 | 0.61 | 2.38 | 0.0 | 0.0 | 0.0 | 0.0 |

Table 7. Frequency of Observed ISL1 Haplotype Pairs In Unrelated Individuals

| HAP1 | HAP2 | Total | CA | AF | AS | HL | AM |
|---|---|---|---|---|---|---|---|
| 3 | 3 | 10.98 | 4.76 | 20.0 | 0.0 | 16.67 | 33.33 |
| 5 | 5 | 28.05 | 38.1 | 10.0 | 45.0 | 11.11 | 66.67 |
| 3 | 13 | 1.22 | 0.0 | 0.0 | 0.0 | 5.56 | 0.0 |
| 5 | 8 | 1.22 | 0.0 | 0.0 | 0.0 | 5.56 | 0.0 |
| 5 | 9 | 1.22 | 0.0 | 0.0 | 5.0 | 0.0 | 0.0 |
| 5 | 14 | 1.22 | 4.76 | 0.0 | 0.0 | 0.0 | 0.0 |
| 5 | 7 | 1.22 | 0.0 | 0.0 | 5.0 | 0.0 | 0.0 |
| 5 | 6 | 1.22 | 0.0 | 0.0 | 0.0 | 5.56 | 0.0 |
| 5 | 1 | 1.22 | 4.76 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 15 | 1.22 | 4.76 | 0.0 | 0.0 | 0.0 | 0.0 |
| 5 | 2 | 1.22 | 4.76 | 0.0 | 0.0 | 0.0 | 0.0 |
| 12 | 4 | 1.22 | 4.76 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 14 | 1.22 | 0.0 | 5.0 | 0.0 | 0.0 | 0.0 |
| 3 | 10 | 1.22 | 0.0 | 5.0 | 0.0 | 0.0 | 0.0 |
| 3 | 12 | 6.1 | 0.0 | 25.0 | 0.0 | 0.0 | 0.0 |
| 5 | 12 | 8.54 | 4.76 | 15.0 | 5.0 | 11.11 | 0.0 |
| 5 | 3 | 30.49 | 28.57 | 15.0 | 40.0 | 44.44 | 0.0 |
| 3 | 11 | 1.22 | 0.0 | 5.0 | 0.0 | 0.0 | 0.0 |

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5% in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from that population group. In addition, the size and composition of the Index Repository means that the relative frequencies determined therein for the haplotypes and haplotype pairs of the ISL1 gene are likely to be similar to the relative frequencies of these ISL1 haplotypes and haplotype pairs in the general U.S. population and in the four population groups represented in the Index Repository. The genetic diversity observed for the three Native Americans is presented because it is of scientific interest, but due to the small sample size it lacks statistical significance.

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and

shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) gene of an individual, which comprises determining which of the ISL1 haplotypes shown in the table immediately below defines one copy of the individual's ISL1 gene, wherein each of the ISL1 haplotypes comprises a set of polymorphisms whose locations and identities are set forth in the table immediately below:

| Haplotype Number[a] | | | | | | | | | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | |
| A | A | A | A | A | A | A | A | G | G | 1 | 3542 |
| A | A | A | A | A | A | A | A | A | A | 2 | 3732 |
| G | G | G | G | G | G | G | G | C | G | 3 | 3789 |
| A | A | G | G | G | G | G | G | G | G | 4 | 3794 |
| A | A | A | A | A | A | A | A | A | A | 5 | 3987 |
| C | C | C | C | C | C | C | T | C | C | 6 | 4078 |
| G | G | G | G | G | G | G | G | G | A | 7 | 4084 |
| C | C | C | C | C | C | G | C | C | C | 8 | 7837 |
| C | G | C | C | C | C | C | C | C | C | 9 | 7885 |
| T | C | C | T | T | T | T | C | C | C | 10 | 8183 |
| C | C | C | C | C | C | T | C | C | C | 11 | 10193 |
| G | G | G | G | G | G | G | G | G | G | 12 | 11502 |
| A | A | A | A | A | G | A | A | A | A | 13 | 13835 |
| G | G | G | A | G | G | G | G | G | G | 14 | 14055 |

| Haplotype Number[a] | | | | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|---|---|---|
| 11 | 12 | 13 | 14 | 15 | | |
| G | G | G | G | G | 1 | 3542 |
| A | A | A | A | G | 2 | 3732 |
| G | G | G | G | G | 3 | 3789 |
| G | G | G | G | G | 4 | 3794 |
| A | A | A | G | A | 5 | 3987 |
| C | C | T | C | C | 6 | 4078 |
| G | G | G | G | G | 7 | 4084 |
| C | C | C | C | C | 8 | 7837 |
| C | C | C | C | C | 9 | 7885 |
| C | C | C | C | C | 10 | 8183 |
| C | C | C | C | C | 11 | 10193 |
| A | G | G | G | G | 12 | 11502 |
| A | A | A | A | A | 13 | 13835 |
| G | G | G | G | G | 14 | 14055 |

[a]Alleles for haplotypes are presented 5′ to 3′ in each column;
[b]PS = polymorphic site;
[c]Position of PS within SEQ ID NO:1.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS14 on the one copy of the individual's ISL1 gene.

3. A method for haplotyping the ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) gene of an individual, which comprises determining which of the ISL1 haplotype pairs shown

in the table immediately below defines both copies of the individual's ISL1 gene, wherein each of the ISL1 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

| Haplotype Pair[a] | | | | | | | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|---|---|---|---|---|---|
| 3/3 | 5/5 | 3/13 | 5/8 | 5/9 | 5/14 | 5/7 | 5/6 | | |
| A/A | A/A | A/G | A/A | A/G | A/G | A/A | A/A | 1 | 3542 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/A | 2 | 3732 |
| G/G | G/G | G/G | G/G | G/C | G/G | G/G | G/G | 3 | 3789 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 4 | 3794 |
| A/A | A/A | A/A | A/A | A/A | A/G | A/A | A/A | 5 | 3987 |
| C/C | C/C | C/T | C/T | C/C | C/C | C/C | C/C | 6 | 4078 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 7 | 4084 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/G | C/C | 8 | 7837 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 9 | 7885 |
| C/C | T/T | C/C | T/C | T/C | T/C | T/T | T/T | 10 | 8183 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/T | C/C | 11 | 10193 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 12 | 11502 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/G | 13 | 13835 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 14 | 14055 |

| Haplotype Pair[a] | | | | | | | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|---|---|---|---|---|---|
| 5/1 | 3/15 | 5/2 | 12/4 | 3/14 | 3/10 | 3/12 | 5/12 | | |
| A/A | A/G | A/A | G/A | A/G | A/G | A/G | A/G | 1 | 3542 |
| A/A | A/G | A/A | A/A | A/A | A/A | A/A | A/A | 2 | 3732 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 3 | 3789 |
| G/A | G/G | G/A | G/G | G/G | G/G | G/G | G/G | 4 | 3794 |
| A/A | A/A | A/A | A/A | A/G | A/A | A/A | A/A | 5 | 3987 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 6 | 4078 |
| G/G | G/G | G/G | G/G | G/G | G/A | G/G | G/G | 7 | 4084 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 8 | 7837 |
| C/C | C/C | C/G | C/C | C/C | C/C | C/C | C/C | 9 | 7885 |
| T/T | C/C | T/C | C/T | C/C | C/C | C/C | T/C | 10 | 8183 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 11 | 10193 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 12 | 11502 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/A | 13 | 13835 |
| G/G | G/G | G/G | G/A | G/G | G/G | G/G | G/G | 14 | 14055 |

| Haplotype Pair[a] | | PS | PS |
|---|---|---|---|
| 5/3 | 3/11 | No.[b] | Pos.[c] |
| A/A | A/G | 1 | 3542 |
| A/A | A/A | 2 | 3732 |
| G/G | G/G | 3 | 3789 |
| G/G | G/G | 4 | 3794 |
| A/A | A/A | 5 | 3987 |
| C/C | C/C | 6 | 4078 |
| G/G | G/G | 7 | 4084 |
| C/C | C/C | 8 | 7837 |
| C/C | C/C | 9 | 7885 |
| T/C | C/C | 10 | 8183 |
| C/C | C/C | 11 | 10193 |
| G/G | G/A | 12 | 11502 |
| A/A | A/A | 13 | 13835 |
| G/G | G/G | 14 | 14055 |

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1.

4. The method of claim 3, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS14 on both copies of the individual's ISL1 gene.

5. A method for genotyping the ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) gene of an individual, comprising determining for the two copies of the ISL1 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13 and PS14, wherein the one or more PS have the location and alternative alleles shown in SEQ ID NO:1.

6. The method of claim 5, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid mixture comprising both copies of the ISL1 gene, or a fragment thereof, that are present in the individual;

(b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

7. The method of claim 5, which comprises determining for the two copies of the ISL1 gene present

in the individual the identity of the nucleotide pair at each of PS1-PS14.

8. A method for haplotyping the ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) gene of an individual which comprises determining, for one copy of the ISL1 gene present in the individual, the identity of the nucleotide at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13 and PS14, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:1.

9. The method of claim 8, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid sample containing only one of the two copies of the ISL1 gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid sample a target region containing the selected polymorphic site;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

10. A method for predicting a haplotype pair for the ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) gene of an individual comprising:

(a) identifying a ISL1 genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13 and PS14, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:74;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) comparing the possible haplotype pairs to the haplotype pair data set forth in the table immediately below; and

(d) assigning a haplotype pair to the individual that is consistent with the data

| Haplotype Pair[a] | | | | | | | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|---|---|---|---|---|---|
| 3/3 | 5/5 | 3/13 | 5/8 | 5/9 | 5/14 | 5/7 | 5/6 | | |
| A/A | A/A | A/G | A/A | A/G | A/G | A/A | A/A | 1 | 3542 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/A | 2 | 3732 |
| G/G | G/G | G/G | G/G | G/C | G/G | G/G | G/G | 3 | 3789 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 4 | 3794 |
| A/A | A/A | A/A | A/A | A/A | A/G | A/A | A/A | 5 | 3987 |
| C/C | C/C | C/T | C/T | C/C | C/C | C/C | C/C | 6 | 4078 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 7 | 4084 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/G | C/C | 8 | 7837 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 9 | 7885 |
| C/C | T/T | C/C | T/C | T/C | T/C | T/T | T/T | 10 | 8183 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/T | C/C | 11 | 10193 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 12 | 11502 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/G | 13 | 13835 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 14 | 14055 |

| Haplotype Pair[a] | | | | | | | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|---|---|---|---|---|---|
| 5/1 | 3/15 | 5/2 | 12/4 | 3/14 | 3/10 | 3/12 | 5/12 | | |
| A/A | A/G | A/A | G/A | A/G | A/G | A/G | A/G | 1 | 3542 |
| A/A | A/G | A/A | A/A | A/A | A/A | A/A | A/A | 2 | 3732 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 3 | 3789 |
| G/A | G/G | G/A | G/G | G/G | G/G | G/G | G/G | 4 | 3794 |
| A/A | A/A | A/A | A/A | A/G | A/A | A/A | A/A | 5 | 3987 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 6 | 4078 |
| G/G | G/G | G/G | G/G | G/G | G/A | G/G | G/G | 7 | 4084 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 8 | 7837 |
| C/C | C/C | C/G | C/C | C/C | C/C | C/C | C/C | 9 | 7885 |
| T/T | C/C | T/C | C/T | C/C | C/C | C/C | T/C | 10 | 8183 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 11 | 10193 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 12 | 11502 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/A | 13 | 13835 |
| G/G | G/G | G/G | G/A | G/G | G/G | G/G | G/G | 14 | 14055 |

| Haplotype Pair[a] | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|
| 5/3 | 3/11 | | |
| A/A | A/G | 1 | 3542 |
| A/A | A/A | 2 | 3732 |
| G/G | G/G | 3 | 3789 |
| G/G | G/G | 4 | 3794 |
| A/A | A/A | 5 | 3987 |
| C/C | C/C | 6 | 4078 |
| G/G | G/G | 7 | 4084 |
| C/C | C/C | 8 | 7837 |
| C/C | C/C | 9 | 7885 |
| T/C | C/C | 10 | 8183 |
| C/C | C/C | 11 | 10193 |
| G/G | G/A | 12 | 11502 |
| A/A | A/A | 13 | 13835 |
| G/G | G/G | 14 | 14055 |

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1.

11. The method of claim 10, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS14, which have the location and alternative alleles shown in SEQ ID NO:1.

12. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-15 shown in the table presented immediately below, wherein each of the haplotypes comprises a set of polymorphisms whose locations and identities are set forth in the table immediately below:

| Haplotype Number[a] | | | | | | | | | | PS | PS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | No.[b] | Pos.[c] |
| A | A | A | A | A | A | A | A | G | G | 1 | 3542 |
| A | A | A | A | A | A | A | A | A | A | 2 | 3732 |
| G | G | G | G | G | G | G | G | C | G | 3 | 3789 |
| A | A | G | G | G | G | G | G | G | G | 4 | 3794 |
| A | A | A | A | A | A | A | A | A | A | 5 | 3987 |
| C | C | C | C | C | C | C | T | C | C | 6 | 4078 |
| G | G | G | G | G | G | G | G | G | A | 7 | 4084 |
| C | C | C | C | C | C | G | C | C | C | 8 | 7837 |
| C | G | C | C | C | C | C | C | C | C | 9 | 7885 |
| T | C | C | T | T | T | T | C | C | C | 10 | 8183 |
| C | C | C | C | C | C | T | C | C | C | 11 | 10193 |
| G | G | G | G | G | G | G | G | G | G | 12 | 11502 |
| A | A | A | A | A | G | A | A | A | A | 13 | 13835 |
| G | G | G | A | G | G | G | G | G | G | 14 | 14055 |

| Haplotype Number[a] | | | | | PS | PS |
|---|---|---|---|---|---|---|
| 11 | 12 | 13 | 14 | 15 | No.[b] | Pos.[c] |
| G | G | G | G | G | 1 | 3542 |
| A | A | A | A | G | 2 | 3732 |
| G | G | G | G | G | 3 | 3789 |
| G | G | G | G | G | 4 | 3794 |
| A | A | A | G | A | 5 | 3987 |
| C | C | T | C | C | 6 | 4078 |
| G | G | G | G | G | 7 | 4084 |
| C | C | C | C | C | 8 | 7837 |
| C | C | C | C | C | 9 | 7885 |
| C | C | C | C | C | 10 | 8183 |
| C | C | C | C | C | 11 | 10193 |
| A | G | G | G | G | 12 | 11502 |
| A | A | A | A | A | 13 | 13835 |
| G | G | G | G | G | 14 | 14055 |

[a]Alleles for haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;

and wherein the haplotype pair is selected from the haplotype pairs shown in the table immediately below, wherein each of the ISL1 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

| Haplotype Pair[a] | | | | | | | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|---|---|---|---|---|---|
| 3/3 | 5/5 | 3/13 | 5/8 | 5/9 | 5/14 | 5/7 | 5/6 | | |
| A/A | A/A | A/G | A/A | A/G | A/G | A/A | A/A | 1 | 3542 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/A | 2 | 3732 |
| G/G | G/G | G/G | G/G | G/C | G/G | G/G | G/G | 3 | 3789 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 4 | 3794 |
| A/A | A/A | A/A | A/A | A/A | A/G | A/A | A/A | 5 | 3987 |
| C/C | C/C | C/T | C/T | C/C | C/C | C/C | C/C | 6 | 4078 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 7 | 4084 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/G | C/C | 8 | 7837 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 9 | 7885 |
| C/C | T/T | C/C | T/C | T/C | T/C | T/T | T/T | 10 | 8183 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/T | C/C | 11 | 10193 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 12 | 11502 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/G | 13 | 13835 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 14 | 14055 |

| Haplotype Pair[a] | | | | | | | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|---|---|---|---|---|---|
| 5/1 | 3/15 | 5/2 | 12/4 | 3/14 | 3/10 | 3/12 | 5/12 | | |
| A/A | A/G | A/A | G/A | A/G | A/G | A/G | A/G | 1 | 3542 |
| A/A | A/G | A/A | A/A | A/A | A/A | A/A | A/A | 2 | 3732 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 3 | 3789 |
| G/A | G/G | G/A | G/G | G/G | G/G | G/G | G/G | 4 | 3794 |
| A/A | A/A | A/A | A/A | A/G | A/A | A/A | A/A | 5 | 3987 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 6 | 4078 |
| G/G | G/G | G/G | G/G | G/G | G/A | G/G | G/G | 7 | 4084 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 8 | 7837 |
| C/C | C/C | C/G | C/C | C/C | C/C | C/C | C/C | 9 | 7885 |
| T/T | C/C | T/C | C/T | C/C | C/C | C/C | T/C | 10 | 8183 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 11 | 10193 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 12 | 11502 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/A | 13 | 13835 |
| G/G | G/G | G/G | G/A | G/G | G/G | G/G | G/G | 14 | 14055 |

| Haplotype Pair[a] | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|
| 5/3 | 3/11 | | |
| A/A | A/G | 1 | 3542 |
| A/A | A/A | 2 | 3732 |
| G/G | G/G | 3 | 3789 |
| G/G | G/G | 4 | 3794 |
| A/A | A/A | 5 | 3987 |
| C/C | C/C | 6 | 4078 |
| G/G | G/G | 7 | 4084 |
| C/C | C/C | 8 | 7837 |
| C/C | C/C | 9 | 7885 |
| T/C | C/C | 10 | 8183 |
| C/C | C/C | 11 | 10193 |
| G/G | G/A | 12 | 11502 |
| A/A | A/A | 13 | 13835 |
| G/G | G/G | 14 | 14055 |

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;

wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

13. The method of claim 12, wherein the trait is a clinical response to a drug targeting ISL1.

14. An isolated genotyping oligonucleotide for detecting a polymorphism in the ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13 and PS14, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:1.

15. The isolated genotyping oligonucleotide of claim 14, which is an allele-specific oligonucleotide that specifically hybridizes to an allele of the ISL1 gene at a region containing the polymorphic site.

16. The allele-specific oligonucleotide of claim 15, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-17, the complements of SEQ ID NOS:4-17, and SEQ ID NOS:18-45.

17. The isolated genotyping oligonucleotide of claim 14, which is a primer-extension oligonucleotide.

18. The primer-extension oligonucleotide of claim 17,which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:46-73.

19. A kit for genotyping the ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) gene of an individual, which comprises a set of oligonucleotides designed to genotype each of polymorphic sites (PS) PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13 and PS14, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:1.

20. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting

of:

(a) a first nucleotide sequence which comprises a ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) isogene, wherein the ISL1 isogene is selected from the group consisting of isogenes 1- 4 and 6 - 15 shown in the table immediately below and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1- 4 and 6 - 15 is further defined by the corresponding set of polymorphisms whose locations and identities are set forth in the table immediately below

| Isogene Number[a] | | | | | | | | | | PS | PS | Region |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | No.[b] | Pos.[c] | Examined[d] |
| A | A | A | A | A | A | A | A | G | G | 1 | 3542 | 3264-4277 |
| A | A | A | A | A | A | A | A | A | A | 2 | 3732 | 3264-4277 |
| G | G | G | G | G | G | G | G | C | G | 3 | 3789 | 3264-4277 |
| A | A | G | G | G | G | G | G | G | G | 4 | 3794 | 3264-4277 |
| A | A | A | A | A | A | A | A | A | A | 5 | 3987 | 3264-4277 |
| C | C | C | C | C | C | C | T | C | C | 6 | 4078 | 3264-4277 |
| G | G | G | G | G | G | G | G | G | A | 7 | 4084 | 3264-4277 |
| C | C | C | C | C | C | G | C | C | C | 8 | 7837 | 7653-8333 |
| C | G | C | C | C | C | C | C | C | C | 9 | 7885 | 7653-8333 |
| T | C | C | T | T | T | T | C | C | C | 10 | 8183 | 7653-8333 |
| C | C | C | C | C | C | T | C | C | C | 11 | 10193 | 9853-10596 |
| G | G | G | G | G | G | G | G | G | G | 12 | 11502 | 11423-11985 |
| A | A | A | A | A | G | A | A | A | A | 13 | 13835 | 13651-14430 |
| G | G | G | A | G | G | G | G | G | G | 14 | 14055 | 13651-14430 |

| Isogene Number[a] | | | | | PS | PS | Region |
|---|---|---|---|---|---|---|---|
| 11 | 12 | 13 | 14 | 15 | No.[b] | Pos.[c] | Examined[d] |
| G | G | G | G | G | 1 | 3542 | 3264-4277 |
| A | A | A | A | G | 2 | 3732 | 3264-4277 |
| G | G | G | G | G | 3 | 3789 | 3264-4277 |
| G | G | G | G | G | 4 | 3794 | 3264-4277 |
| A | A | A | G | A | 5 | 3987 | 3264-4277 |
| C | C | T | C | C | 6 | 4078 | 3264-4277 |
| G | G | G | G | G | 7 | 4084 | 3264-4277 |
| C | C | C | C | C | 8 | 7837 | 7653-8333 |
| C | C | C | C | C | 9 | 7885 | 7653-8333 |
| C | C | C | C | C | 10 | 8183 | 7653-8333 |
| C | C | C | C | C | 11 | 10193 | 9853-10596 |
| A | G | G | G | G | 12 | 11502 | 11423-11985 |
| A | A | A | A | A | 13 | 13835 | 13651-14430 |
| G | G | G | G | G | 14 | 14055 | 13651-14430 |

[a]Alleles for isogenes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;
[d]Region examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1;

(b) a second nucleotide sequence which comprises a fragment of the first nucleotide sequence, wherein the fragment comprises one or more polymorphisms selected from the group

consisting of guanine at PS1, guanine at PS2, cytosine at PS3, adenine at PS4, guanine at PS5, thymine at PS6, adenine at PS7, guanine at PS8, guanine at PS9, cytosine at PS10, thymine at PS11, adenine at PS12, guanine at PS13 and adenine at PS14, wherein the selected polymorphism has the location set forth in the table immediately above; and

(c) a third nucleotide sequence which is complementary to the first or second nucleotide sequence.

21. The isolated polynucleotide of claim 20, which is a DNA molecule and comprises both the first and third nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

22. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 20, wherein the organism expresses a ISL1 protein encoded by the first nucleotide sequence.

23. The recombinant nonhuman organism of claim 22, which is a transgenic animal.

24. The isolated polynucleotide of claim 20 which consists of the second nucleotide sequence.

25. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a coding sequence for a ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) isogene wherein the coding sequence is defined by a haplotype selected from the group consisting of 2c and 7c shown in the table immediately below and wherein the coding sequence comprises SEQ ID NO:2 except at each of the polymorphic sites which have the locations and polymorphisms set forth in the table immediately below:

| Coding Sequence Haplotype Number[a] | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|
| 2c | 7c | | |
| G | C | 9 | 243 |
| C | T | 11 | 655 |

[a]Alleles for coding sequence haplotypes are presented 5′ to 3′ in each column; the numerical portion of the coding sequence haplotype number represents the number of the parent full ISL1 haplotype;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:2;

and

(b) a fragment of the coding sequence, wherein the fragment comprises at least one polymorphism selected from the group consisting of guanine at a position corresponding to nucleotide 243 and thymine at a position corresponding to nucleotide 655, wherein said positions in the coding sequence and the fragment refer to SEQ ID NO:2.

26. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 25, wherein the organism expresses a ISL1 transcription factor, LIM/homeodomain,

(islet-1) (ISL1) protein encoded by the polymorphic variant sequence.

27. The recombinant nonhuman organism of claim 26, which is a transgenic animal.

28. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:3 and the polymorphic variant comprises one or more variant amino acids selected from the group consisting of methionine at a position corresponding to amino acid position 81 and serine at a position corresponding to amino acid position 219.

29. An isolated monoclonal antibody specific for and immunoreactive with the isolated polypeptide of claim 28.

30. A method for screening for drugs targeting the isolated polypeptide of claim 28 which comprises contacting the ISL1 polymorphic variant with a candidate agent and assaying for binding activity.

31. A computer system for storing and analyzing polymorphism data for the ISL1 transcription factor, LIM/homeodomain, (islet-1) gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the haplotypes set forth in the table immediately below:

| Haplotype Number[a] | | | | | | | | | | PS | PS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | No.[b] | Pos.[c] |
| A | A | A | A | A | A | A | A | G | G | 1 | 3542 |
| A | A | A | A | A | A | A | A | A | A | 2 | 3732 |
| G | G | G | G | G | G | G | G | C | G | 3 | 3789 |
| A | A | G | G | G | G | G | G | G | G | 4 | 3794 |
| A | A | A | A | A | A | A | A | A | A | 5 | 3987 |
| C | C | C | C | C | C | C | T | C | C | 6 | 4078 |
| G | G | G | G | G | G | G | G | G | A | 7 | 4084 |
| C | C | C | C | C | C | G | C | C | C | 8 | 7837 |
| C | G | C | C | C | C | C | C | C | C | 9 | 7885 |
| T | C | C | T | T | T | T | C | C | C | 10 | 8183 |
| C | C | C | C | C | C | T | C | C | C | 11 | 10193 |
| G | G | G | G | G | G | G | G | G | G | 12 | 11502 |
| A | A | A | A | A | G | A | A | A | A | 13 | 13835 |
| G | G | G | A | G | G | G | G | G | G | 14 | 14055 |

| Haplotype Number[a] | | | | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|---|---|---|
| 11 | 12 | 13 | 14 | 15 | | |
| G | G | G | G | G | 1 | 3542 |
| A | A | A | A | G | 2 | 3732 |
| G | G | G | G | G | 3 | 3789 |
| G | G | G | G | G | 4 | 3794 |
| A | A | A | G | A | 5 | 3987 |
| C | C | T | C | C | 6 | 4078 |
| G | G | G | G | G | 7 | 4084 |
| C | C | C | C | C | 8 | 7837 |
| C | C | C | C | C | 9 | 7885 |
| C | C | C | C | C | 10 | 8183 |
| C | C | C | C | C | 11 | 10193 |
| A | G | G | G | G | 12 | 11502 |
| A | A | A | A | A | 13 | 13835 |
| G | G | G | G | G | 14 | 14055 |

[a]Alleles for haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;

and the haplotype pairs set forth in the table immediately below:

| Haplotype Pair[a] | | | | | | | | PS No.[b] | PS Pos.[c] |
|---|---|---|---|---|---|---|---|---|---|
| 3/3 | 5/5 | 3/13 | 5/8 | 5/9 | 5/14 | 5/7 | 5/6 | | |
| A/A | A/A | A/G | A/A | A/G | A/G | A/A | A/A | 1 | 3542 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/A | 2 | 3732 |
| G/G | G/G | G/G | G/G | G/C | G/G | G/G | G/G | 3 | 3789 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 4 | 3794 |
| A/A | A/A | A/A | A/A | A/A | A/G | A/A | A/A | 5 | 3987 |
| C/C | C/C | C/T | C/T | C/C | C/C | C/C | C/C | 6 | 4078 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 7 | 4084 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/G | C/C | 8 | 7837 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 9 | 7885 |
| C/C | T/T | C/C | T/C | T/C | T/C | T/T | T/T | 10 | 8183 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/T | C/C | 11 | 10193 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 12 | 11502 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/G | 13 | 13835 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 14 | 14055 |

| Haplotype Pair[a] | | | | | | | | PS | PS |
|---|---|---|---|---|---|---|---|---|---|
| 5/1 | 3/15 | 5/2 | 12/4 | 3/14 | 3/10 | 3/12 | 5/12 | No.[b] | Pos.[c] |
| A/A | A/G | A/A | G/A | A/G | A/G | A/G | A/G | 1 | 3542 |
| A/A | A/G | A/A | A/A | A/A | A/A | A/A | A/A | 2 | 3732 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 3 | 3789 |
| G/A | G/G | G/A | G/G | G/G | G/G | G/G | G/G | 4 | 3794 |
| A/A | A/A | A/A | A/A | A/G | A/A | A/A | A/A | 5 | 3987 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 6 | 4078 |
| G/G | G/G | G/G | G/G | G/G | G/A | G/G | G/G | 7 | 4084 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 8 | 7837 |
| C/C | C/C | C/G | C/C | C/C | C/C | C/C | C/C | 9 | 7885 |
| T/T | C/C | T/C | C/T | C/C | C/C | C/C | T/C | 10 | 8183 |
| C/C | C/C | C/C | C/C | C/C | C/C | C/C | C/C | 11 | 10193 |
| G/G | G/G | G/G | G/G | G/G | G/G | G/G | G/G | 12 | 11502 |
| A/A | A/A | A/A | A/A | A/A | A/A | A/A | A/A | 13 | 13835 |
| G/G | G/G | G/G | G/A | G/G | G/G | G/G | G/G | 14 | 14055 |

| Haplotype Pair[a] | | PS | PS |
|---|---|---|---|
| 5/3 | 3/11 | No.[b] | Pos.[c] |
| A/A | A/G | 1 | 3542 |
| A/A | A/A | 2 | 3732 |
| G/G | G/G | 3 | 3789 |
| G/G | G/G | 4 | 3794 |
| A/A | A/A | 5 | 3987 |
| C/C | C/C | 6 | 4078 |
| G/G | G/G | 7 | 4084 |
| C/C | C/C | 8 | 7837 |
| C/C | C/C | 9 | 7885 |
| T/C | C/C | 10 | 8183 |
| C/C | C/C | 11 | 10193 |
| G/G | G/A | 12 | 11502 |
| A/A | A/A | 13 | 13835 |
| G/G | G/G | 14 | 14055 |

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1.

32. A genome anthology for the ISL1 transcription factor, LIM/homeodomain, (islet-1) (ISL1) gene which comprises ISL1 isogenes defined by any one of haplotypes 1-15 set forth in the table shown below:

| Haplotype Number[a] | | | | | | | | | | PS | PS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | No.[b] | Pos.[c] |
| A | A | A | A | A | A | A | A | G | G | 1 | 3542 |
| A | A | A | A | A | A | A | A | A | A | 2 | 3732 |
| G | G | G | G | G | G | G | G | C | G | 3 | 3789 |
| A | A | G | G | G | G | G | G | G | G | 4 | 3794 |
| A | A | A | A | A | A | A | A | A | A | 5 | 3987 |
| C | C | C | C | C | C | C | T | C | C | 6 | 4078 |
| G | G | G | G | G | G | G | G | G | A | 7 | 4084 |
| C | C | C | C | C | C | G | C | C | C | 8 | 7837 |
| C | G | C | C | C | C | C | C | C | C | 9 | 7885 |
| T | C | C | T | T | T | T | C | C | C | 10 | 8183 |
| C | C | C | C | C | C | T | C | C | C | 11 | 10193 |
| G | G | G | G | G | G | G | G | G | G | 12 | 11502 |
| A | A | A | A | A | G | A | A | A | A | 13 | 13835 |
| G | G | G | A | G | G | G | G | G | G | 14 | 14055 |

| Haplotype Number[a] | | | | | PS | PS |
|---|---|---|---|---|---|---|
| 11 | 12 | 13 | 14 | 15 | No.[b] | Pos.[c] |
| G | G | G | G | G | 1 | 3542 |
| A | A | A | A | G | 2 | 3732 |
| G | G | G | G | G | 3 | 3789 |
| G | G | G | G | G | 4 | 3794 |
| A | A | A | G | A | 5 | 3987 |
| C | C | T | C | C | 6 | 4078 |
| G | G | G | G | G | 7 | 4084 |
| C | C | C | C | C | 8 | 7837 |
| C | C | C | C | C | 9 | 7885 |
| C | C | C | C | C | 10 | 8183 |
| C | C | C | C | C | 11 | 10193 |
| A | G | G | G | G | 12 | 11502 |
| A | A | A | A | A | 13 | 13835 |
| G | G | G | G | G | 14 | 14055 |

[a]Alleles for haplotypes are presented 5′ to 3′ in each column

[b]PS = polymorphic site;

[c]Position of PS in SEQ ID NO:1.

**POLYMORPHISMS IN THE ISL1 GENE**

TCATACCCAC TGGTATCTCT TCCTTTTCTC TGAAATTAGA GCACAGATTC
TCCTGATTGT TCCTCAATAA GCCACCACCA CCAACGGGAA GAGACGCAAA 100
TTGGTTCTAA ACACCCACAC CCCGTTCGCG GTGGAATATA AATGGTTAAT
CGTAACTCCC TTTAAACGCA CATAATTTTA TCTGACCCTT GGACATGATT 200
AAAGATACTG AGGATTTATA TAGCCTTATT CTACATCATT AGAGACTACA
CAGAAAAGAT AAAAACGGGA AAGGGGATAC ATACGCATCT ACTCAATATC 300
CATTTTTAGA GGTAAATAAA ATCTTCCTGT TTCATTTTGG CCCCCAAATT
ATATGATGTG GACATCAGAA AATATAAATC CTTCAGTTAT ACATATTTTT 400
TCATCTGAAT ACCTTACTCA GTAAAACTGG CAGTGTAACT TATTCTACTC
TTTCAAACCT GATTCTGATG TGAGCTAAAC ATATAGCATC ATCTCCCTGG 500
GAATAAATTC TCAAATTTAT ACCTCCATTG GAAAGCTGAT CAAAGCCCCT
CTTCCCTTCC CAGTCTTGGT TAGATGAGAA GGAATAAGCT GTCTTACTGT 600
TTGTGATTTC CCAAAGAAAC AACTTAAAAT TCATTTATAT CTGCCAGAGG
TTATTGGGAG ATAAATAAAT GCTAAATGTG TAATTTTACC TTTCAATTGC 700
AATTAAAACC TATTTAGATT TTTGAAATGT AGCTTGAGAG CATGATGTTT
GTGCAGGAAT CTGTATCTGA CATTAAGTTT GATGTGGAAT GTTTAAATGG 800
TAACTAAAGC AGAGAAAGCA ACAGTACACA AAAACACAAG TCTAAGAGCC
CATCAGGAGG TTGGGTGTAA CAGCCACATT TTGGAAGGTG TATACAAGGG 900
CACATGGGCA GATGCCACAG ACTCCAGAGG AACTTAAATT TCAGTATCCC
TTTTATATTG AAAGACACTG TCAGATATAT TGAGGCAGCC TGGAGGTGAT 1000
GGGGACTACA CTTAATAACT ACTCATACTC CTCATTTTCT CAGTAAAGGA
ATCAAACCAG AGATTCACAC TATTGCCGTT TTCAAGGATA TATTTTTGAC 1100
CTAACAATAT ACAAAGGTCA AAACGAGTCC TTTTTTGAGA GGAGAGGGTC
TAATCCCCGC TTTGAACACA AGAGGTCGCT CCTGTGATCA AGTATTACGG 1200
ATAAATTGCT CATCAGTTTT ATGTATACAA AATTTTACAG ATATATTTGA
TGCACTAAAT CCACATACAG CAATCAGCTG AGGAAATCAT AATAATCCAT 1300
GATCAATTTA GAACTTCAAG GTAAAGTAAA ACTACAAAGC CTGCTATTCT
AACTTGGATT TTAATGCAAA ATGTGTCATG ATACATGCAC TGAACTTCAG 1400
AGCATGTCCA TCATGGTGGG GTCCCTCAAA CGTCTCCTTT TAATGCTGGT
ATCTTCACCA TGCGTCTTGC TTCTATCAGT TTAGGAACTT ATTAAGAAAA 1500
TACTTGTTTA CAACAATGGT GCAGAATGAA AAGAGGAAAA TGTTGTCTTC
ACCATTAAGG TGAAATTGTC TATTACTAAA GGTTTCCCCC AAGCCGCATG 1600
CAAGCTTTTC TACTTGGAAG TGCTATCATC TGCAATAATT CTTTGTAATA
GCTTATTGGT TTAGGGAGTT AAAATACCAC ACTGTGGGTG GGAGGTGGGG 1700
GCTGGTTTGA AGGAGGATGG ATAGTGAGGG TGCTTTGTGG ATCTCAAATG
TAAGTCATCA GGCTTCTGGT GGCTTCTGGT GTTTAATTCC ATCTTATAAA 1800
AGTGGCATGA TGGAATATGA CTAGGTAAAA CTCTATTTTC TGGCATTGTG
GTGGTATCTT TTTTTTTTTT TTTTTTTACT AGGGTCAACA TGTTTATCTC 1900
TATGGGAAAC ACACACAAAT ACTAGGAAGA TTTGCTTCGT ATTGAGATTA
ACAGCATCAG CTTTGCAAAA GCAGCCTGAC AAAAGGCTCC ATACATCTAA 2000
TACCAGAGAC TAGAAACAGA CCAAAAAAAA AAAAAAAAAA AAACACACAC
CACACAACTG GTAAACTTAT CAACCTAGGT CTATCCAACT TTCGCTTTCA 2100
TACACGCGCT CTAAAACAAC TTATAATAAA AATGGAAGGG AAGACAGATC
CAATTTGAAA TCCTCCTTGA GAAAAAACAA ATCAAAACTA GTTCCTGGGG 2200
AAGAAAGCCT CAGCTAGGTC AGGAGGAAAC TTACGCATCT TTGATTCCCT
TTCCGCTTTG GAGAGGCATT ACCCGTTCAA GCCCAGCCAA TGCGGCCGGC 2300
CAGGATTTAC AGCTCTTATC AAGGGTTAGA TTTTGCGAAA GATATAAAGA
AAGGAGTTTC CTAACACACC GGGTCTTCTC ACAGCAACAA GACACCGAAA 2400
TTAAGCCTTC TATGGTTTGT GTCTGTAGGA CTTTTTAGAT AAAACTGGCA
GCCCCATTAG ATAAGAAATG GCTTTTTAGA AGCTTTGGGG GAAGCGCGGG 2500
GGTTCTTCGT GGCCTTCCTG TGACTGTCTT TGGGAGACCG TAACAGATGA
TGGGAACATG TAAGAATGAT TGAAGAAGGT TACCGGCGAG TTGCACCTAA 2600

FIGURE 1A

```
CAGTCTTCTC TCTCCAAGTA GTTTCTCGAG AGAAAATTAT ATATTAACGC
CTGGGGACGT GAAATAGGTC TTTCACTGTT CTGTAATAAA TCCCTCCCTA      2700
CTCGCGCCCT TAGACTAATG CATTTTTAAA AAATCTCCAC AAAGTCATGG
AGACTTGTCC TTATAATTAC CAAAATACCA CAGCTAGTTG CACAAGTTTG      2800
CTTAAAAACA AAGGGCAGGA GACCAGCAGT GAGAAGAATT GAACAGTATA
TAAAAATCTT ACAAACTTTA GACTACTGAA AGATCATTTC GCTCTTTGGC      2900
TTCGGGTTAG TGAGCTTCAA TTCTAGAGTT AAACTGAGAA GCAGACCTCG
CGGGTCTGAA AGACAAAAAG TCAGTCCGCG GAGTCAACTT CGGTTTACCG      3000
GTGCGTTCCA ACCTGATTGT CCCAACAGCA AGAAGCGCCC CTTTCCTCCC
ACCCAACGTT TTTAGATATC TGAGGTTGGG GGAGACGGAG TAATGGTGTA      3100
GCAGTCAGGA AACAGTTCCT CTGGGTTTAA TCATCCCATT CCCGATTCTT
CTTCCCCTAG CGCGCCGAGC GCAGCCAGTT CTGGAGGAAC TAGCCCTCCG      3200
GAGCAGCCAA GGCAGGCCAG GCCCCGGGGA TGTGCACCAC GGCGTGGGGA
CCAAACCAAG CTGAACGGCC TGCTCCAAGT TCCCCCCTTT TCCTAGAGCC      3300
CGGGAGGGAC CTCGTCAACA GGTAGACCTA CCCACCAGCC ACTCGCCACC
GGCGGCAGAA AGTGAGCCCC GCGCGCAACG CGGCCAGCCG GACTGCGGGG      3400
ACCCCAGGAG CGCAGGGCGG AGGAGCAGCG CCACAGGAGG CGAGGGCGCA
GGCGGCGCGG CCGGGAAGGA ACGCGGGAGG GGACAGAAGG AAGAGGAAGA      3500
GGAGGAGAGG GAGGCCAGAG CCAGAACAGC CCGGCAGCCC GAGCTTCGGG
                                                    G
GGAGAACGGC CTGAGCCCCG AGCAAGTTGC CTCGGGAGCC CTAATCCTCT      3600
CCCGCTGGCT CGCCGAGCGG TCAGTGGCGC TCAGCGGCGG CGAGGCTGAA
ATATGATAAT CAGAACAGCT GCGCCGCGCG CCCTGCAGCC AATGGGCGCG      3700
GCGCTCGCCT GACGTCCCCG CGCGCTGCGT CAGACCAATG GCGATGGAGC
                                  G
TGAGTTGGAG CAGAGAAGTT TGAGTAAGAG ATAAGGAAGA GAGGTGCCCG      3800
                                            C     A
AGCCGCGCCG AGTCTGCCGC CGCCGCAGCG CCTCCGCTCC GCCAACTCCG
CCGGCTTAAA TTGGACTCCT AGATCCGCGA GGGCGCGGCG CAGCCGAGCA      3900
GCGGCTCTTT CAGCATTGGC AACCCCAGGG GCCAATATTT CCCACTTAGC
CACAGCTCCA GCATCCTCTC TGTGGGCTGT TCACCAACTG TACAACCACC      4000
                                        G
ATTTCACTGT GGACATTACT CCCTCTTACA GATATGGGAG ACATGGGAGA
[exon 1: 4043..
TCCACCAAAA AGTAAGAGGC TATTTTACCT TGTGGGGCTC GGTGTGCTGT      4100
                           T      A
          ..4061]
TCTTGTGCGG GGTTCTCTCT CAGGCACAGG CTGAGGTGCC AAGGGCTCTT
TGGAGTTGGA GTCATTGCCT GGAGAAAGAG AAAAGGTGGC TTTTTCTTGT      4200
TGCCGCCACG CCTGCATGCT TACTGTCGGT TCTTATCTTC GGGAAACTGA
TTGTACCTTG TGTGTGAATT CGCCTGTGTG CCCTCCAAAG CTCTAGCTTT      4300
CTGGTGCTAA GCGGTGATTT CCTCCTGGGG AATCCTGAGC TCTCCGAGAA
GGTTATTATG TTGCAAAGGT CTGCCTGCAC AGTCAATGCC CAGAGATGTG      4400
AATTAGCATT AGACTTGCAA AAGAGAACGA GTGACAACTG TATTTATGCC
TGCTCTTGCT AACAATATCC AGTCCTGTGT GCTATTTAAG AGCGCGCTTC      4500
ACGGAAAATA TAGACATCCC TGCGTTCACT TAACGCTTCT AGTCAAAACC
TTTTCTTTGA CTTGACTTAT CCATAATCTT TCCCAATGAT TATAGCAAAG      4600
AGGAAGGGGG GGGGGAGAAA TACAAAATGA GCGGGTTTGA TTGCGTGCTA
GGCGTACAAA TGTAGACTAT TCCAATCTGC ATTTTACATA TATTCCACCT      4700
CCTTTTAAAA ATGAGTCAAG GTTTTGATGG CACATTTCAA TTACCATCCC
AAAGTGCAAT GCTCTAAAAA AAAAAAAAAA GAAAGAAAGA AAGAAAGAAA      4800
AAAACCTCCC AGAGTACGCC CTATAAGAGA ACGACACTAA AAGTGTGTTT
ATCTCTGTAG GAAGTAAACG GTTAGTCAAT CATGTATTTA TTTTCATTTC      4900
```

FIGURE 1B

```
AGAAAAACGT CTGATTTCCC TATGTGTTGG TTGCGGCAAT CAGATTCACG
[exon 2: 4903..
ATCAGTATAT TCTGAGGGTT TCTCCGGATT TGGAATGGCA TGCGGCATGT      5000
TTGAAATGTG CGGAGTGTAA TCAGTATTTG GACGAGAGCT GTACATGCTT
TGTTAGGGAT GGGAAAACCT ACTGTAAAAG AGATTATATC AGGTATGGCA      5100
          ..5092]
TTTACACTTC TTTCTTAATT TTGTGGGATT TCCCTGAATC TCCCCACTCT
TTATGTATTA TTTGGTGTGG CTTTGTCTTT TTGTGAAGTT TGCCTCAGTG      5200
TAGTCATACA AGCCAAAGTT ACCCTGTACA TGTGTTAAAA AAATCAAGCT
ATGCTGTTCA TTTCATTCTT TAGTTGAGAA AAACAAAAAC CCTTAACAGT      5300
GGTATTCATA ATTCCGGGGT ATTGAGGCTT GTTTAATTAC TCTTGGAGTT
TATGATGCAC AAATTATTTT CCTCTTTCAC CCTCCCCCTT ACAAAACAAA      5400
ATTTTAAAAA GATGGAGAAG TTTGGATTTT TAGCTTTAAA ATAGGGTTGA
TTTTTGTTGT ATAGTGCAGT GTTCTGTTTG TTTTAGTCCT TTTTAAAATT      5500
AGTAGCTTAC AAATTCTTTG GTGGCATCAA TGCAATAGGT GAAATAAAAG
TTTGACCGAA GCATGTTTAG AGATGTACTT TGAAAGAGCG AGTACAGGTA      5600
TTGCTCCTTT TATTTTTGGG GTAAGACCTC CTTCTGAGAA AAATTTAAAA
CCAACCTAAA TATTCCTTGG AAAAAACACC GGAAACTTAA TCTTTTTAAA      5700
TATTAACCCT TTGGTGACAT CTAACTGTCT CTTCTTTCTT ATCTTATCTG
AGCTGATGAA TTAGAGCAGA TCAAATTGCC CATCATCTGT CTACGAACAA      5800
TTGGTATATT TAGATAATTG AACAGCTTCC TTTCTCACAT TAAAATCTGG
TAACTGATAA AATGAGCGAA TTGTCCAAAT TGACAAGACT GAAACAACAT      5900
AGGAACTTTC TGAGTTTGGT TTTGTTGTTT TGGAGAGTTT TTGTTTTTTT
TTTCCTCCAA TTTATTCTGC AACACGTTTT GCTAATCTCA AGTTTCCTCT      6000
GACTTGTGTG TATGTATCAG AAACTTTGTT TTCTGCCTTA GAAAGCCAGT
AGTCTCTAAA GAAAATTGTA TTCATTTTAT TAACAAACAG AAGAGACATC      6100
AGCATCATTA TTATGTTAAA TAATAGCAAA ATATCACTTT TTAAATGTCC
GGTGGCTATT AACAAGTAAT TAATTAGCTT TTGTTAGGCA AATGGTTTCT      6200
GGAGCTTGAG AACTTTTATT AAAGTTTAGT TAAGATTTAA TATACAGTCA
CAGTTTGCTC CTGCTCACTT AGTATCCAGC ATTTTTTTCT TCTTTTTTAA      6300
AAATCATGAC ACAGAGAGTA TAATCTTGGT AGATAAAATT AACCTGGTTG
GGGGAGGTTA ATACTTCGGA GAGGGAGTGA AAGGAAGTAA GGGAAGTCGG      6400
GGTACAGGAA GGGGGAGGGA TTTTCTAAAT TGTTTGGTCA CCGCCAAAGT
CAAGTCTTCA CCCTATGAAA TGGAAGATCT CACATTGAGT AGGCGGAGGG      6500
AGGAAAAACT TTTGAGTCCA CCTTCTAACC TCTGACAAAT GAGCGTTTTC
ATTGTTTACT AGATTGGTGT GTAAACGCAA GATTCTAGAG AAGGAGAGCC      6600
CACTTCAGGA GTATCTTTAC TGCTATGGAA ATAGTATTTT GCTCAATTGC
ACACAGGCTT GCATGTGCCT AATTCTGGAT ACACACATGT GTAGAAGGAA      6700
CTAATCATTT TTACCTTCTC TTCACTCTCT CTCAACTCTG TGTGTGTGTG
TGTGTGTGTG TGTGTGTGTG TGTGTAATCT TGTAGTTGTA AAAGCAGAAC      6800
AGACTGGACA GTTAGATTTC CACATCTCTC CTTGGAGAAG CAGGATGCCT
CCTCCTGTTA TGTGGATCTT TTCCTCTCTC TTCCATTCTT TCTGTTCGTA      6900
GGAATGCCCC AGCTTCTGTT ATTCCTGAAA GATGGAGAAG GGGCCAGGGA
AGTGCAGCCT AGATGGAACC TATAAAGATT GTCCCTTGGT AAGGAAAGGC      7000
CAGGAGTGAG AAAGACCTTA GAAGCGGGTC TTTGCATTTT TTTCATTCTG
GTCATGGTTT TCAAGAAAAT TGAAATGAGG TAGATGATTC AGCAACTTGA      7100
AAAAGATTGA GGGAACAGAC GCAGATTTTT TTAAAAAAAT AATAATACAA
GGAAGAATGG AGAGGAAATT TTCTGTTAAC ATTGCTGCCT GAAGAAAATC      7200
TTTAGTTGGA GAAAGACTGG AAAGTACTTG TGCAAAAGGA GATGTGGAAA
CTCTCAGAGG TTTCATTTTG TTATTCTGCT TGTTTATTTG TGAGTGTTTG      7300
CAAACCGAGT GGGGTGACAA TCCCCTTCTC CTACCTCCTT TTTTCTTGGA
AGGAGGACTT TTTGTTGCAG TTTTAGACAT TTCTAGCAGC AGAAATTGTG      7400
GGATAGGGAA GTGAAAGTGT TGGTGTCGGT GGCCACCAGA GTCTTTCTGG
ATTCCTTCCT GCCAAGATCT GCAAGATCAA CACTGGGATT GATTGCTAGA      7500
```

FIGURE 1C

```
GCAGCAGCCC GAGTTTGGAA CCCATCAATA CATTTTCTGT GGTACAAGCT
AGGTGTTTTG AGCTAAGAGT TACCAACTAA GACAGAGGTT CATCGGAAAG      7600
GAAACGGGAG TAAAAGAAAG GGAGGAGGGA GGGAGGGGAA AAGAGAGATG
GGGGAAGGAA GAGAGACAGG GAAGGAGAGA GCAGGGTTTC ATTTCTGTCC      7700
TTCTGTTTCC AACTTCTGTT TGGAAATGCT GTTTACTTGG GGCGTCTTGC
CCGGGATCTT GGGCCAGGGA AGTGCCGGCC TGAAGTGACC CCCTCTTCCT      7800
GTACTTCTCT CCCCGCTCTG GGCCGCCTCC GCTCCCCCCT CCCCCGCACA
                                          G
GGTTGTACGG GATCAAATGC GCCAAGTGCA GCATCGGCTT CAGCAAGAAC      7900
                                       G
[exon 3: 7852..
GACTTCGTGA TGCGTGCCCG CTCCAAGGTG TATCACATCG AGTGTTTCCG
CTGTGTGGCC TGCAGCCGCC AGCTCATCCC TGGGGACGAA TTTGCGCTTC      8000
GGGAGGACGG TCTCTTCTGC CGAGCAGACC ACGATGTGGT GGAGAGGGCC
AGTCTAGGCG CTGGCGACCC GCTCAGTCCC CTGCATCCAG CGCGGCCACT      8100
GCAAATGGCA GGTACTCCTC TGCCCGGCTC GGGTAGGCAG GCGCCAGGTT
          ..8111]
AAGCCAGCCT GTGTGCCAGC GGCCACAACA ACTATGGTAG CTACAGGGGT      8200
                                       C
GGTCGTAGTG TTTGCCTGCA GTTAAATGAA GTGTTCTGTA TGCAATTTGC
GCTGTGCTCT GCTCCTTTGC AGCAAGGTTC AATGCACTCA CTGTCTCCCT      8300
TGATTCCCCG AGCACACCTA CACCGTCTGT GTGTCTCTAT ATGGTTACAC
ATAAATGTAC ACCACTTGTG TACACGTGTA TACACACGCC CAAACATTAC      8400
TTCCAGTTCG CTCTGGCCTC CAAACCTTGG CTTGCTGAAA ACGGGCTTCA
GCTCCCAGCC AGGTATTCTC CTGCTGCCTA ATTAAAGGGG CGGAGCCCCG      8500
GGTCCCTGGA GCTTCATCCT TTAACCCAAT GAAGGAAGCT TAGGTGGCCT
GAAGTCATTT AGTCTCCCAA ATCCTTTTTC CTTGTGAGTT GCTTCACACT      8600
CGAAATTTTT TTTTTAATTT TTTTATCTTT CTGTGAGAGA ACAGGACTGA
AAAGATACAG TTTTAAAAAC TGCAGGCCAT TGCACAGAGT TGTAATATAA      8700
AACTGTCAAC AAGCTTATCT GCAGTAATTG CCTTTTAAAG GGAGCCTGCT
TCTTTAAATC ATTCATTCTA TATGATTTGG TGAGAATTTC ATCTTCAGGC      8800
CCATGGTTGT AGCTCTAAAT TGACCCCATA GGTGTTGGCC TGACCCTAGG
GGGTTGTAGA AGGTGCAGGA TTTGTATCAT GTAGATAAGA GGACTCATTC      8900
CCAAGGAAGA GGAGTGGAAA CACAGCAAGG TTGGCCGGGA CCAAAGCAGT
GGGTTAGAAG GTGGACAGTG TTTCCAAACC TGACTTCCTG CCATGAATAG      9000
ATCTACCCCT TTGCAGTTTT AAAGTATCAA TTCCCACTAA ACACTGAAGG
TGAGGAAACT ATAGCCCTCC CTTACCCTTC TGCCTTCTGG CAGCTCTAAG      9100
AATTCTGTTC AGGGGGATTT GTGACTAGTT TGCACCGGGG CACGGCTGGG
GTGGTGCTCC TGTTCAGTGG AGCCTGCACT CTGCTTGTGG GGAAGCACAG      9200
AGGAAGCTAA AATACCGAGA GGGAGGCGGG GGACATCTCC CAGCCACCGT
TTATCTAGAG CCTAGGCAGC TCAACAGAGT TTCCGTTTTC CACTGCTTGG      9300
GATCAGCCCA TCTCAGGAAC ATCCATGTAT TACCTTAGAT TTAATACTAA
GAGCAGGGAT TGGAGATATG GCAGAAATAG CGAATCTCTT CAGCCCCTTC      9400
ACATGACTGT CCTCTCGGAC TGAAGTTCAA GGCGTTCTGG CAGAGTTCTC
GACCTTCCCC TTGCAGAAGT CCCTGCTGGT GTAGTATTTA TGGCTGTCAC      9500
TGAAGTGCTC TGCGTTCCTT TCCCTGGTAC CCTCTGTGGC CTTGGCCCAA
GAGAAAATTC TGATCCTGGA GAGGGTGGTA ATCAATGTAA CTGGGGCCCA      9600
GTCTGGGCAC AAGGAAAGGT GAGAATGGAG GAGAAACAGT GCTGAAAAAT
GCCACCCCTG CTGTGAACAG GGGGACAGAC TTTGAGACCT GCTTCCCTTG      9700
GCTAACACTT TGTTGACACG AGGAGGGGCG AGTGCTGCGT TTCAGGCCGG
GATTACTCAG CAAAGACCTC TGCAGATTAG AGAGGAAGAT TTTATTCTCC      9800
CTTTCACCCT CTTCGCCCCC ACCTCTGCCG CCCCCTGCTT TGTGTGCTGA
GGCTGCAAAC CCTAGCCATT GTCCTGAGTA TCTCGGGCGG GCGAGCAAGT      9900
AAGCGGGCGG GCGGGCGGGC AAGCGAGCGA GCGAGCGAGC GCGCGACCGC
```

FIGURE 1D

```
GGGCGGGCCG GCAAGCGAGC CTCCAGCCCA GCGCTCACGG CGCTCCTTGC      10000
CCCGCAGCGG AGCCCATCTC CGCCAGGCAG CCAGCCCTGC GGCCCCACGT
[exon 4: 10008..
CCACAAGCAG CCGGAGAAGA CCACCCGCGT GCGGACTGTG CTGAACGAGA      10100
AGCAGCTGCA CACCTTGCGG ACCTGCTACG CCGCAAACCC GCGGCCAGAT
GCGCTCATGA AGGAGCAACT GGTAGAGATG ACGGGCCTCA GTCCCCGTGT      10200
                                                    T
GATCCGGGTC TGGTTTCAAA ACAAGCGGTG CAAGGACAAG AAGCGAAGCA
TCATGATGAA GCAACTCCAG CAGCAGCAGC CCAATGACAA AACTGTGAGT      10300
          ..10294]
GGCTCTGGGG CCGGGCAGGG AATGCGAGGG GGAAGGAGAC GCAGCGTGCG
AGGTGCGTTC CTGGTACGCA GGATCGCACG GTTTTCAATC CTGCTCCTGG      10400
GCAGGAGTTT GGCCGGGGCT GCCCCTCATC CTTACCCCCC TACCCATGCC
CCGGGGGACA GGCTACCCGG CGCCGGCCGC CAGCTGAGGG CGGGGAAGCT      10500
GGGAGGCTCC GTGCGCCGGG GGAGCAGCAT CCAGGTCCCA ACCTCGTGGG
TGGGCTCATG CCCTTCCACC TCGCCTGTAC CTGTGAACCG GAGAAACGCC      10600
GTCCTCCCCT CTGAGGGCAG GCGGCAACGA GGTTTGGCCC GGGTTTTGCC
AACATTCAGA TCGTCAGTTC CTCACGTACA CAAGAAGAGG GAGGGATAAT      10700
ACCTTGGATT CCTGCCTACA TCCAGGGGTT CCGTGGGCAG GTCACCCTGT
GAGCCCCCAG GGCGCACCGC ACTTCTAAGT AAGGTCGGCC GCTGCGCCTT      10800
CAGGCTGGCG AGTTCCCCCA AGGTGACCCG CATGCCCAGA TCACCCTCTG
CTCCAGGTGA AGCCCAGGCC TCCACAGAGG CATCAGGCCC CTCGCACCAG      10900
TATCCACTGT TATCTTGGTC CCACGGAAGC ACCCACTCTG CAGGCCTCCT
GGTGAAGTTA AGCTAGAGTT TCTTTTCTTC CTTTTTTTCT TTTCTTTCTT      11000
TTTCTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTACTGCTTT
GGACCTATTT TTAAATGCCA TAAAATCTGC TGTCATTAAA CTTGGCAGGC      11100
TGGCCAAGAT TGGGCCAGGG CACTTTCTGA GTTGGTTAGT GCATAATAGC
ACAATAGGAA CCAGACCCAA ATGCTTTGGG GGGATGGAGT GGGGGGCTGG      11200
CTCTTCCTTG AGGAGAACGG CTTGGAAAAA ATCTGCAGCT AACTGAAACT
GCTCAGAAAA CCACCCTGTC TAGAGGCTGA AGGGAAGCCC TGCTTACCTC      11300
AGCTTTTTAG TTCTGGGAAG CTATGGTCTG AGAAGGCAGA GGGGAGGAAT
TGGGCTGAGC TGTGAAGGTA AGGGGGAAGA AGAAAATCAA AGTAGAATTT      11400
GGTTTAATAA GGTCCATGCA GACCTAATAG TCCAGCCCAC AGAGGCAGAA
AAACAAAACA ATAAAACAAA TTGAATTCTA ACTAATATCC GTAGGTACGG      11500
CGGATTAACT GAGTCAATAA AGACCACTAT ATAGATAAGA TAATACCAGG
 A
GTATATTTGC TTAGCCTGTG CAGACAACGG AGGGAGGGAA TTTGCTCATT      11600
AACATGTTGG GATTGGTTGG GGGGCCTATT CACAGAATAT CCAGGGGATG
[exon 5: 11636..
ACAGGAACTC CCATGGTGGC TGCCAGTCCA GAGAGACACG ACGGTGGCTT      11700
ACAGGCTAAC CCAGTGGAAG TACAAAGTTA CCAGCCACCT TGGAAAGTAC
TGAGCGACTT CGCCTTGCAG AGTGACATAG ATCAGCCTGC TTTTCAGCAA      11800
CTGGTAAGTG TCAGCTCCCA GATGGAAGAG GCTGAATTCC CAACAGGAGA
          ..11803]
CTCTGGTTTA ACTGTCACAC ATTGAAAGAT TCAGTGGGGA GGGTGCCTTC      11900
TTGGGCTCAG GGTTGGGGAG AAACCAAGGA GGTGGGTAAT GAAGAGAAGG
GAGACAAATG CAGGGAAAAC GAACCTCTTG GCATCTTTTT TTTTTTAATG      12000
AGACTGCATA ATTTGACCAT ATAGGTTGAA TTTTCTATCA ATCAGGCCTT
CTTTGAAGGA TTAATTTCAA GGTACCTAAC TCTAGGTAGC ATGTGCCAGA      12100
AGATGTACAG TGTTGGAGAA TCATACATCT TAGAATTTTA GAGTTGTCAA
GGACTTCAGG AAATCGTCTT GGCATTTCAA TCAGCAATTA GTAAGTTTAT      12200
CCTTCCTGAG CATCTAGAGA ATGGGATATA TAGGACCCAA ATCAAGGCGA
TTGTAGTATA TAATCAATGC TATAATACAC AATCAGTGTT GTTATCAATA      12300
AACAGCAGGC ATGTGTCTGG GTACAATTTT CAAATATATT AATAAAGATT
```

FIGURE 1E

```
ATTCAAATAG ATGAATACTT TTTGTTACAG TATCCCTTGC TGGGAATGTC      12400
TTAATCTAAA ATGTAGGACC GTTTAAATGT TTTCAAGTGT ATGAGTTCAA
ATGTCATAGA GACACACAGT GTGTACCATG TATAGCAAAA GGGACATGAG      12500
CTCTACCAAT CAGAAGTAAA GTGTTTAATC TGTGAAAACC TTAACATGTT
TTCCACATCC AGAGAGGAGA AAATTAATTC ACTTTTTGCC TACAAAAGGC      12600
TTAAGGGGTC AAGATAAATA AGAACAATAA ATATATGTCC TTTGTAATAT
GCTATATTTA TATAGATGAT TTTTTTTTCT TAAAGAGTAA TCAGCCTTAT      12700
AGAATCTTGT TTTATAAAAT GTAAAGATCT ATCCTGAAAC CTTGTTCCCT
TTTTTGGAAA TGAAGCTTTA GTTGAGGTTA GCTTTTTACC CTCATATTTA      12800
CCTGGAGGGC ATTTGCTTTC TCAATGTCAA CAGTTAGGTA ATTGGCCAGA
GGCAAGTGGT TAAAAGGGCT TGGCCCCAGG CTTGTGTTTG CAAATGCTAA      12900
GTGGGTGCAG AGGCTAGAAG TCCCTTAATC TCATATTGGA AAAATTTACT
GTAGAAAGAA ATGTAGGCTC TAGAACTAGG AAAAAAAAAT TATTCTAAGC      13000
TCATTAATCT GTTGAGTTAT TTGAGCGAAT CCTGAATCAC AGGAGGAAGG
TAAGGGGAGG CTTCAGGGCA GCCAAATGTT TGCACTTTCT GAAACTTTAG      13100
TGTCAGATGA GAGCAGTGGA AGGGAAGCTG AGGCAGGAGT GGGCATAGTT
AGAGAAGGTT TACAACAGCA GTACAATGCG TTTAGGGTTA AAAGAAGGAG      13200
TCAGATATTT AAGAAGGAGT CAGATATTAG GGTTAAAAGA AGGAGTCAGA
ATGGGATGAT GTCATAATAT ATGGGTCTCA TTTTGGAAGG AAGAGCCTGA      13300
TTTAAAGAGA GAGAGAGAGA AAGGCCAAGA GGCAGCAGGA CCAACAAGGA
AGAATGCCCA AGCTGTGAGC CTGCTGAGGA GTTAATCTTT GTTCTGTGGA      13400
GCCTCCTCTC AATCTCCTGT CAAAGGATCT GAGCCTGTTA CGGATTTTCC
AACTGAAGAA GAGAGTCTTT GATGCCTAGA GACTGAGAGC TCACCTACTC      13500
CCAGGGCAAC ATGTAGCCAG CAGGATAATT TTATTTCGAG CATGCATAGT
AGAGTTGTGA TGCCATTTTA CAGTGGGAAA CACATTTGTT CTTAAATAAT      13600
TTAATGCAAC ATAATGTTGG GAATTCAGTT TCAGTTAAAA CAGAGATCTT
TTGGAAGATG GGAAAGTGAG AGGATTTCTT CCCAAGTTTT TCTCCTCTAG      13700
GCTTTCTCTA AGCCTGTTAA AATTCAGTTA TCTATGTGAA TATCTTTACA
TATCTATCTA CACAAACATT TCTACATATA CAATATGATG AGTTTATAAT      13800
CTTTTTATGA ATACTATTCC AGTGTCCTTT ATTTATTTCT CAACCTTCTA-
                                  G
TGCAGGTCAA TTTTTCAGAA GGAGGACCGG GCTCTAATTC CACTGGCAGT      13900
[exon 6: 13856..
GAAGTAGCAT CAATGTCCTC TCAACTTCCA GATACACCTA ACAGCATGGT
AGCCAGTCCT ATTGAGGCAT GAGGAACATT CATTCTGTAT TTTTTTTCCC      14000
          ..13972]
TGTTGGAGAA AGTGGGAAAT TATAATGTCG AACTCTGAAA CAAAAGTATT
TAACGACCCA GTCAATGAAA ACTGAATCAA GAAATGAATG CTCCATGAAA      14100
    A
TGCACGAAGT CTGTTTTAAT GACAAGGTGA TATGGTAGCA ACACTGTGAA
GACAATCATG GGATTTTACT AGAATTAAAC AACAAACAAA ACGCAAAACC      14200
CAGTATATGC TATTCAATGA TCTTAGAAGT ACTGAAAAAA AAAGACGTTT
TTAAAACGTA GAGGATTTAT ATTCAAGGAT CTCAAAGAAA GCATTTTCAT      14300
TTCACTGCAC ATCTAGAGAA AAACAAAAAT AGAAAATTTT CTAGTCCATC
CTAATCTGAA TGGTGCTGTT TCTATATTGG TCATTGCCTT GCCAAACAGG      14400
AGCTCCAGCA AAAGCGCAGG AAGAGAGACT GGCCTCCTTG GCTGAAAGAG
TCCTTTCAGG AAGGTGGAGC TGCATTGGTT TGATATGTTT AAAGTTGACT      14500
TTAACAAGGG GTTAATTGAA ATCCTGGGTC TCTTGGCCTG TCCTGTAGCT
GGTTTATTTT TTACTTTGCC CCCTCCCCAC TTTTTTTGAG ATCCATCCTT      14600
TATCAAGAAG TCTGAAGCGA CTATAAAGGT TTTTGAATTC AGATTTAAAA
ACCAACTTAT AAAGCATTGC AACAAGGTTA CCTCTATTTT GCCACAAGCG      14700
TCTCGGGATT GTGTTTGACT TGTGTCTGTC CAAGAACTTT TCCCCCAAAG
ATGTGTATAG TTATTGGTTA AAATGACTGT TTTCTCTCTC TATGGAAATA      14800
AAAAGGAAAA AAAAAAAGGA AACTTTTTTT GTTTGCTCTT GCATTGCAAA
```

FIGURE 1F

```
AATTATAAAG TAATTTATTA TTTATTGTCG GAAGACTTGC CACTTTTCAT      14900
GTCATTTGAC ATTTTTTGTT TGCTGAAGTG AAAAAAAAAG ATAAAGGTTG
TACGGTGGTC TTTGAATTAT ATGTCTAATT CTATGTGTTT TGTCTTTTTC      15000
TTAAATATTA TGTGAAATCA AAGCGCCATA TGTAGAATTA TATCTTCAGG
ACTATTTCAC TAATAAACAT TTGGCATAGA TAAATAAATA AACACAGTGA      15100
TCCAGGGTCC CTCTTTTATG TTTACAAACA AGCTATTCAA ATGTCTGAAA
GTAAGGATTG CATGTGAATT GGGGTAAACT GACACACAGT AGGGGTGATA      15200
TTGTATTGCA AGTAGTTAAG GTCGAGTTTG CATGTTTGGG TTTCAGTAGT
TGGATCAAAG TTGGTTGGAA GCCCCTTAGC TCTCAGCTAT GATAATGCTA      15300
CTGGTTTTGT GAAAATGAGA TGCTGTTCTC TGTCCTGCCC CTCTCTTTCC
TTCCCCCAGT CTCTTTCTTG CCATACCCCA CCCATCCACC CACCCCAAAT      15400
AGGAAGAGGG TCTTGCTTAA AATGGCTTTT GATTCAGAGT CACTATTTTA
TTTTCTTAGA ATACGTGATT TATACCAAAG TCTCCTGCTT GGGATTGTTG      15500
AAATGACCCA TATGTCAAAT ATATAAGCTC TTGGTAAGAG AGACTTTCTG
CCTAGCCAGT GAGGTATGAC CATTATTTAC CAGTGGCGGC TGTCAGCCGA      15600
TTGAGAAATG CTGCTCTATT TGTGCAGAAG GAACAGGCCT AGAGAAGAGA
GTATGCTCTG CAATTCCCAG ACTAGTCGGT GTATGCATTT TCTTGCCTCT      15700
TCTCCCCTCT CCCACCTCCG CTTCGTCCCT TGCCCAAGTC CAATTTCAAA
AATCTTGGGA AGCATGACTG GCTATGGAGA GAAACATGGC ATATATACTG      15800
TTGGGAAACC ACCTTTTATG GGATGGGGTA GCAGGTCATG TGCTTTGCTG
GACCCATCAG CTATCAGGGT TACGGGCATG CTCTCTGGGA CAGACTATTC      15900
CTCCCCTGGC ATTTAGCAAT AGTTCAAGTG CAAGAGCCAG AACCCTGGAA
AAGATAAATG GCACTGCCCG AAATCGGGCT TCTTGGTGCT TCTTCATTTG      16000
CTGGCATTGA ATTAGGATTT GGGGCTCCAC CCTGTTCCCC CTCCTCCCTT
TACCCCACCC CAGGATGATG CTTTGACTCT GGAAACTGCA TTGATTTGTA      16100
TGCGCGAGTG TAATGCCATT TTCAGTACCT GACAATTCTT ATTATGAAAG
CTAAAGTACC CTAATTGTGT GTCATGTGGC GTACCCCTCC AGGCTCGATA      16200
GGGTGAGGTG GAGGCTGTGC CCCCCAGCTC TCAAGTCACT TTCACCCTCT
GCAAACCTGC CTCAAGTCTG GAAGCCAGCC TCAATTAGGC TTTTGTTTGG      16300
CTTGGTTCTA CATAACGGGT TAATTACAGG CTCTAACGTG GGGGCAGAGA
GAAAACAAAC TTTGCATTGA AAATGTTACA GTGCATTAGA CATGCCCGTT      16400
GATTACTTTG ATTGTTGTGT CTAATTCTTG ACAGAGGGGT AATAATTTGG
CTGGTGGCAA TAATGGCTTG CAGTCTTCTG CTGAGAAAAA CAGAAGTTAT      16500
GTCTTATCTC ACACGGATCT TCTCCGGGAA AGGTTTTTTT GTTGTTGTTT
GTTTGTTTTG TTTTGTTTTG TTTTTCACTC GGATCTCCCT CCTCTAATCT      16600
CTTGCCAAGG GACTTTACTC TTTTAGCTTC CCATTCACTA TTTCTATTGG
TCTCTTTTTT CCTTTTTCCT CACCATCATA GCTGAAATGG CAACTAATAT      16700
GCATGTGTAA GGATGGCAGG TTCAACATGA ATATTGGTGT CATGCCTGCA
AAGTTACTGA GAGTCCAAGC CTTCCTTACT TGGAAAACAC TGGACCCACT      16800
TCCCCAGCCG TAGCCTCTGG TGGACTCAAA AATGTTAGGG ACTAGAGATG
GAAAACACTT TATTTAGGAT TTGAATGCAA TAGAGAGACT TAAGTGAGGC      16900
ACCCATCTTT GAGAATCTTT TTCCTCCTCG AATTGATGAA TTGATTTCTA
AGATTAAGAA TATGTTTGCA TGTTTGGGTG GGGGGCGTTG ATGTTTTACA      17000
TACATAGATG TCAGTATTCT GAAGTCCATC GCTTGAATTT AAGCACAATA
ATTTAGGGAG GGGGTGATGA ATGATTTGGT TTCACTTGCT AGATTCAAGA      17100
TATTTGTTGC CTGATGGCTG AGAATGGATT GATGGGTCAA TAGCATTCTA
GTTCTTTTAT TATTTATACG AAGGTGTGTC TTCCCCTTTT AGACTCCTCA      17200
GCCTGCACTC TAAAATGTGT AGCTGAAATG TCATTTTGTT TGCAAATCAT
CTCTTGCATA CTTTTCTCTG CACACCTCTG GAGACTGAGT TTTGAGCTCT      17300
GCTCTTGTAA TGAAATTCCT CTTGTTCGTT ATTATCTTTA ATCGTCCCTC
CCTTATTACT CTGAATTCTT CCTTCTGGAC ATTATGACTC ATTTCCTCCA      17400
GCTTCTGCAA GCTAAAATTT AATGCAGCTA AAAAACATCT AGTATGCTCC
CCTTAAATCA GAAAATTCAT ATCAATATAC ATAGTCTTAA AGTCAGGATA      17500
TTGTTTCCCC TGTAGTTTCA GCGCAATGCT ATCTTCCATT TCAGCCTATT
```

FIGURE 1G

```
TTCCCTCTAT CCCATTCATC AAATGAACTT GGGTTTTCTT ACAAATCAAA      17600
AAAGTTCTTT TCTAAGTTAG AAGCATGTTT GAAGATTTTT TTCTCCTTTG
AAACAAGGTT TCTGGTTATT TGTAATTCTG TGTGTACCCA CTTGCCCATT      17700
GACGACTTGC TTGGAGTTCA CGATTAAATA GATGACTTTG TGATGCTTGG
CTGCTATTCA CCCTGATCAA CTTGTTAAAC TCATAATGTT GACCTCTTTT      17800
CTGGTGTATC TTATCCTTTT GAGACTAGCT AAAAAGACAA TGTGCTTTCA
GGTTACGTTG GAAGTACCTT GTATTTCCCT AGCACCTTGA ATTTTAAGCC      17900
ATAAGATCTA TCTCATGAAA TGTAACTGTT GTAAAATCCC ATGTGGCTTT
ACTTAGCGTT TTCTTTTTTC T                                     17971
```

FIGURE 1H

**POLYMORPHISMS IN THE CODING SEQUENCE OF ISL1**

```
ATGGGAGATC CACCAAAAAA AAAACGTCTG ATTTCCCTAT GTGTTGGTTG
CGGCAATCAG ATTCACGATC AGTATATTCT GAGGGTTTCT CCGGATTTGG      100
AATGGCATGC GGCATGTTTG AAATGTGCGG AGTGTAATCA GTATTTGGAC
GAGAGCTGTA CATGCTTTGT TAGGGATGGG AAAACCTACT GTAAAAGAGA      200
TTATATCAGG TTGTACGGGA TCAAATGCGC CAAGTGCAGC ATCGGCTTCA
                                                  G
GCAAGAACGA CTTCGTGATG CGTGCCCGCT CCAAGGTGTA TCACATCGAG      300
TGTTTCCGCT GTGTGGCCTG CAGCCGCCAG CTCATCCCTG GGGACGAATT
TGCGCTTCGG GAGGACGGTC TCTTCTGCCG AGCAGACCAC GATGTGGTGG      400
AGAGGGCCAG TCTAGGCGCT GGCGACCCGC TCAGTCCCCT GCATCCAGCG
CGGCCACTGC AAATGGCAGC GGAGCCCATC TCCGCCAGGC AGCCAGCCCT      500
GCGGCCCCAC GTCCACAAGC AGCCGGAGAA GACCACCCGC GTGCGGACTG
TGCTGAACGA GAAGCAGCTG CACACCTTGC GGACCTGCTA CGCCGCAAAC      600
CCGCGGCCAG ATGCGCTCAT GAAGGAGCAA CTGGTAGAGA TGACGGGCCT
CAGTCCCCGT GTGATCCGGG TCTGGTTTCA AAACAAGCGG TGCAAGGACA      700
   T
AGAAGCGAAG CATCATGATG AAGCAACTCC AGCAGCAGCA GCCCAATGAC
AAAACTAATA TCCAGGGGAT GACAGGAACT CCCATGGTGG CTGCCAGTCC      800
AGAGAGACAC GACGGTGGCT TACAGGCTAA CCCAGTGGAA GTACAAAGTT
ACCAGCCACC TTGGAAAGTA CTGAGCGACT TCGCCTTGCA GAGTGACATA      900
GATCAGCCTG CTTTTCAGCA ACTGGTCAAT TTTTCAGAAG GAGGACCGGG
CTCTAATTCC ACTGGCAGTG AAGTAGCATC AATGTCCTCT CAACTTCCAG     1000
ATACACCTAA CAGCATGGTA GCCAGTCCTA TTGAGGCATG A              1041
```

FIGURE 2

**ISOFORMS OF THE ISL1 PROTEIN**

```
MGDPPKKKRL ISLCVGCGNQ IHDQYILRVS PDLEWHAACL KCAECNQYLD
ESCTCFVRDG KTYCKRDYIR LYGIKCAKCS IGFSKNDFVM RARSKVYHIE      100
                                  M
CFRCVACSRQ LIPGDEFALR EDGLFCRADH DVVERASLGA GDPLSPLHPA
RPLQMAAEPI SARQPALRPH VHKQPEKTTR VRTVLNEKQL HTLRTCYAAN      200
PRPDALMKEQ LVEMTGLSPR VIRVWFQNKR CKDKKRSIMM KQLQQQQPND
                    S
KTNIQGMTGT PMVAASPERH DGGLQANPVE VQSYQPPWKV LSDFALQSDI      300
DQPAFQQLVN FSEGGPGSNS TGSEVASMSS QLPDTPNSMV ASPIEA          346
```

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Kliem, Stefanie E.
Koshy, Beena
Tanguay, Debra A.

<120> HAPLOTYPES OF THE ISL1 GENE

<130> MWH-1022PCT ISL1

<140> tba
<141> 2001-08-06

<150> 60/223,535
<151> 2000-08-04

<160> 74

<170> PatentIn Ver. 2.1

<210> 1
<211> 17971
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> (3542)
<223> PS1: polymorphic base A or G

<220>
<221> allele
<222> (3732)
<223> PS2: polymorphic base A or G

<220>
<221> allele
<222> (3789)
<223> PS3: polymorphic base G or C

<220>
<221> allele
<222> (3794)
<223> PS4: polymorphic base G or A

<220>
<221> allele

<222> (3987)
<223> PS5: polymorphic base A or G

<220>
<221> allele
<222> (4078)
<223> PS6: polymorphic base C or T

<220>
<221> allele
<222> (4084)
<223> PS7: polymorphic base G or A

<220>
<221> allele
<222> (7837)
<223> PS8: polymorphic base C or G

<220>
<221> allele
<222> (7885)
<223> PS9: polymorphic base C or G

<220>
<221> allele
<222> (8183)
<223> PS10: polymorphic base T or C

<220>
<221> allele
<222> (10193)
<223> PS11: polymorphic base C or T

<220>
<221> allele
<222> (11502)
<223> PS12: polymorphic base G or A

<220>
<221> allele
<222> (13835)
<223> PS13: polymorphic base A or G

<220>
<221> allele
<222> (14055)
<223> PS14: polymorphic base G or A

<400> 1

tcatacccac tggtatctct tccttttctc tgaaattaga gcacagattc tcctgattgt 60
tcctcaataa gccaccacca ccaacgggaa gagacgcaaa ttggttctaa acacccacac 120
cccgttcgcg gtggaatata aatggttaat cgtaactccc tttaaacgca cataatttta 180
tctgaccctt ggacatgatt aaagatactg aggatttata tagccttatt ctacatcatt 240
agagactaca cagaaaagat aaaaacggga aaggggatac atacgcatct actcaatatc 300
catttttaga ggtaaataaa atcttcctgt ttcattttgg cccccaaatt atatgatgtg 360
gacatcagaa aatataaatc cttcagttat acatattttt tcatctgaat accttactca 420
gtaaaactgg cagtgtaact tattctactc tttcaaacct gattctgatg tgagctaaac 480
atatagcatc atctccctgg gaataaattc tcaaatttat acctccattg gaaagctgat 540
caaagcccct cttcccttcc cagtcttggt tagatgagaa ggaataagct gtcttactgt 600
ttgtgatttc ccaaagaaac aacttaaaat tcatttatat ctgccagagg ttattgggag 660
ataaataaat gctaaatgtg taattttacc tttcaattgc aattaaaacc tatttagatt 720
tttgaaatgt agcttgagag catgatgttt gtgcaggaat ctgtatctga cattaagttt 780
gatgtggaat gtttaaatgg taactaaagc agagaaagca acagtacaca aaaacacaag 840
tctaagagcc catcaggagg ttgggtgtaa cagccacatt ttggaaggtg tatacaaggg 900
cacatgggca gatgccacag actccagagg aacttaaatt tcagtatccc ttttatattg 960
aaagacactg tcagatatat tgaggcagcc tggaggtgat ggggactaca cttaataact 1020
actcatactc ctcatttttct cagtaaagga atcaaaccag agattcacac tattgccgtt 1080
ttcaaggata tatttttgac ctaacaatat acaaaggtca aaacgagtcc ttttttgaga 1140
ggagagggtc taatccccgc tttgaacaca agaggtcgct cctgtgatca agtattacgg 1200
ataaattgct catcagtttt atgtatacaa aattttacag atatatttga tgcactaaat 1260
ccacatacag caatcagctg aggaaatcat aataatccat gatcaattta gaacttcaag 1320
gtaaagtaaa actacaaagc ctgctattct aacttggatt ttaatgcaaa atgtgtcatg 1380
atacatgcac tgaacttcag agcatgtcca tcatggtggg gtccctcaaa cgtctccttt 1440
taatgctggt atcttcacca tgcgtcttgc ttctatcagt ttaggaactt attaagaaaa 1500
tacttgttta caacaatggt gcagaatgaa aagaggaaaa tgttgtcttc accattaagg 1560
tgaaattgtc tattactaaa ggtttccccc aagccgcatg caagcttttc tacttggaag 1620
tgctatcatc tgcaataatt ctttgtaata gcttattggt ttagggagtt aaaataccac 1680
actgtgggtg ggaggtgggg gctggtttga aggaggatgg atagtgaggg tgctttgtgg 1740
atctcaaatg taagtcatca ggcttctggt ggcttctggt gtttaattcc atcttataaa 1800
agtggcatga tggaatatga ctaggtaaaa ctctattttc tggcattgtg gtggtatctt 1860
ttttttttt ttttttttact agggtcaaca tgtttatctc tatgggaaac acacacaaat 1920
actaggaaga tttgcttcgt attgagatta acagcatcag ctttgcaaaa gcagcctgac 1980
aaaaggctcc atacatctaa taccagagac tagaaacaga ccaaaaaaaa aaaaaaaaaa 2040
aaacacacac cacacaactg gtaaacttat caacctaggt ctatccaact ttcgctttca 2100
tacacgcgct ctaaaacaac ttataataaa aatggaaggg aagacagatc caatttgaaa 2160
tcctccttga gaaaaaacaa atcaaaacta gttcctgggg aagaaagcct cagctaggtc 2220
aggaggaaac ttacgcatct ttgattccct ttccgctttg gagaggcatt acccgttcaa 2280
gcccagccaa tgcggccggc caggatttac agctcttatc aagggttaga ttttgcgaaa 2340
gatataaaga aaggagtttc ctaacacacc gggtcttctc acagcaacaa gacaccgaaa 2400
ttaagccttc tatggtttgt gtctgtagga ctttttagat aaaactggca gccccattag 2460
ataagaaatg gcttttttaga agctttgggg gaagcgcggg ggttcttcgt ggccttcctg 2520
tgactgtctt tgggagaccg taacagatga tgggaacatg taagaatgat tgaagaaggt 2580
taccggcgag ttgcacctaa cagtcttctc tctccaagta gtttctcgag agaaaattat 2640
atattaacgc ctggggacgt gaaataggtc tttcactgtt ctgtaataaa tccctcccta 2700
ctcgcgccct tagactaatg catttttaaa aaatctccac aaagtcatgg agacttgtcc 2760
ttataattac caaaatacca cagctagttg cacaagtttg cttaaaaaca aagggcagga 2820

gaccagcagt gagaagaatt gaacagtata taaaaatctt acaaacttta gactactgaa 2880
agatcatttc gctctttggc ttcgggttag tgagcttcaa ttctagagtt aaactgagaa 2940
gcagacctcg cgggtctgaa agacaaaaag tcagtccgcg gagtcaactt cggtttaccg 3000
gtgcgttcca acctgattgt cccaacagca agaagcgccc ctttcctccc acccaacgtt 3060
tttagatatc tgaggttggg ggagacggag taatggtgta gcagtcagga aacagttcct 3120
ctgggtttaa tcatcccatt cccgattctt cttcccctag cgcgccgagc gcagccagtt 3180
ctggaggaac tagccctccg gagcagccaa ggcaggccag gccccgggga tgtgcaccac 3240
ggcgtgggga ccaaaccaag ctgaacggcc tgctccaagt tccccccttt tcctagagcc 3300
cgggagggac ctcgtcaaca ggtagaccta cccaccagcc actcgccacc ggcggcagaa 3360
agtgagcccc gcgcgcaacg cggccagccg gactgcgggg accccaggag cgcagggcgg 3420
aggagcagcg ccacaggagg cgagggcgca ggcggcgcgg ccgggaagga acgcgggagg 3480
ggacagaagg aagaggaaga ggaggagagg gaggccagag ccagaacagc ccggcagccc 3540
grgcttcggg ggagaacggc ctgagccccg agcaagttgc ctcgggagcc ctaatcctct 3600
cccgctggct cgccgagcgg tcagtggcgc tcagcggcgg cgaggctgaa atatgataat 3660
cagaacagct gcgccgcgcg ccctgcagcc aatgggcgcg gcgctcgcct gacgtccccg 3720
cgcgctgcgt crgaccaatg gcgatggagc tgagttggag cagagaagtt tgagtaagag 3780
ataaggaasa gagrtgcccg agccgcgccg agtctgccgc cgccgcagcg cctccgctcc 3840
gccaactccg ccggcttaaa ttggactcct agatccgcga gggcgcggcg cagccgagca 3900
gcggctcttt cagcattggc aaccccaggg gccaatattt cccacttagc cacagctcca 3960
gcatcctctc tgtgggctgt tcaccarctg tacaaccacc atttcactgt ggacattact 4020
ccctcttaca gatatgggag acatgggaga tccaccaaaa agtaagaggc tattttayct 4080
tgtrgggctc ggtgtgctgt tcttgtgcgg ggttctctct caggcacagg ctgaggtgcc 4140
aagggctctt tggagttgga gtcattgcct ggagaaagag aaaaggtggc tttttcttgt 4200
tgccgccacg cctgcatgct tactgtcggt tcttatcttc gggaaactga ttgtaccttg 4260
tgtgtgaatt cgcctgtgtg ccctccaaag ctctagcttt ctggtgctaa gcggtgattt 4320
cctcctgggg aatcctgagc tctccgagaa ggttattatg ttgcaaaggt ctgcctgcac 4380
agtcaatgcc cagagatgtg aattagcatt agacttgcaa aagagaacga gtgacaactg 4440
tatttatgcc tgctcttgct aacaatatcc agtcctgtgt gctatttaag agcgcgcttc 4500
acggaaaata tagacatccc tgcgttcact taacgcttct agtcaaaacc ttttctttga 4560
cttgacttat ccataatctt tcccaatgat tatagcaaag aggaaggggg gggggagaaa 4620
tacaaaatga gcgggtttga ttgcgtgcta ggcgtacaaa tgtagactat tccaatctgc 4680
attttacata tattccacct cctttaaaa atgagtcaag gttttgatgg cacatttcaa 4740
ttaccatccc aaagtgcaat gctctaaaaa aaaaaaaaaa gaaagaaaga aagaaagaaa 4800
aaaacctccc agagtacgcc ctataagaga acgacactaa aagtgtgttt atctctgtag 4860
gaagtaaacg gttagtcaat catgtattta ttttcatttc agaaaaacgt ctgatttccc 4920
tatgtgttgg ttgcggcaat cagattcacg atcagtatat tctgagggtt tctccggatt 4980
tggaatggca tgcggcatgt ttgaaatgtg cggagtgtaa tcagtatttg gacgagagct 5040
gtacatgctt tgttagggat gggaaaacct actgtaaaag agattatatc aggtatggca 5100
tttacacttc tttcttaatt ttgtgggatt tccctgaatc tccccactct ttatgtatta 5160
tttggtgtgg ctttgtcttt ttgtgaagtt tgcctcagtg tagtcataca agccaaagtt 5220
accctgtaca tgtgttaaaa aaatcaagct atgctgttca tttcattctt tagttgagaa 5280
aaacaaaaac ccttaacagt ggtattcata attccggggt attgaggctt gtttaattac 5340
tcttggagtt tatgatgcac aaattatttt cctctttcac cctcccccct acaaaacaaa 5400
attttaaaaa gatggagaag tttggatttt tagctttaaa atagggttga tttttgttgt 5460
atagtgcagt gttctgtttg ttttagtcct ttttaaaatt agtagcttac aaattctttg 5520
gtggcatcaa tgcaataggt gaaataaaag tttgaccgaa gcatgtttag agatgtactt 5580
tgaaagagcg agtacaggta ttgctccttt tatttttggg gtaagacctc cttctgagaa 5640
aaatttaaaa ccaacctaaa tattccttgg aaaaaacacc ggaaacttaa tctttttaaa 5700

tattaaccct ttggtgacat ctaactgtct cttctttctt atcttatctg agctgatgaa 5760
ttagagcaga tcaaattgcc catcatctgt ctacgaacaa ttggtatatt tagataattg 5820
aacagcttcc tttctcacat taaaatctgg taactgataa aatgagcgaa ttgtccaaat 5880
tgacaagact gaaacaacat aggaactttc tgagtttggt tttgttgttt tggagagttt 5940
ttgtttttt tttcctccaa tttattctgc aacacgtttt gctaatctca agtttcctct 6000
gacttgtgtg tatgtatcag aaactttgtt ttctgcctta gaaagccagt agtctctaaa 6060
gaaaattgta ttcatttat taacaaacag aagagacatc agcatcatta ttatgttaaa 6120
taatagcaaa atatcacttt ttaaatgtcc ggtggctatt aacaagtaat taattagctt 6180
ttgttaggca aatggtttct ggagcttgag aacttttatt aaagtttagt taagatttaa 6240
tatacagtca cagtttgctc ctgctcactt agtatccagc attttttttct tctttttaa 6300
aaatcatgac acagagagta taatcttggt agataaaatt aacctggttg ggggaggtta 6360
atacttcgga gagggagtga aaggaagtaa gggaagtcgg ggtacaggaa gggggaggga 6420
ttttctaaat tgtttggtca ccgccaaagt caagtcttca ccctatgaaa tggaagatct 6480
cacattgagt aggcggaggg aggaaaaact tttgagtcca ccttctaacc tctgacaaat 6540
gagcgttttc attgtttact agattggtgt gtaaacgcaa gattctagag aaggagagcc 6600
cacttcagga gtatctttac tgctatggaa atagtatttt gctcaattgc acacaggctt 6660
gcatgtgcct aattctggat acacacatgt gtagaaggaa ctaatcattt ttaccttctc 6720
ttcactctct ctcaactctg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtaatct 6780
tgtagttgta aaagcagaac agactggaca gttagatttc cacatctctc cttggagaag 6840
caggatgcct cctcctgtta tgtggatctt ttcctctctc ttccattctt tctgttcgta 6900
ggaatgcccc agcttctgtt attcctgaaa gatggagaag gggccaggga agtgcagcct 6960
agatggaacc tataaagatt gtcccttggt aaggaaaggc caggagtgag aaagacctta 7020
gaagcgggtc tttgcatttt tttcattctg gtcatggttt tcaagaaaat tgaaatgagg 7080
tagatgattc agcaacttga aaaagattga gggaacagac gcagattttt ttaaaaaaat 7140
aataatacaa ggaagaatgg agaggaaatt ttctgttaac attgctgcct gaagaaaatc 7200
tttagttgga gaaagactgg aaagtacttg tgcaaaagga gatgtggaaa ctctcagagg 7260
tttcattttg ttattctgct tgtttatttg tgagtgtttg caaaccgagt ggggtgacaa 7320
tccccttctc ctacctcctt ttttcttgga aggaggactt tttgttgcag ttttagacat 7380
ttctagcagc agaaattgtg ggatagggaa gtgaaagtgt tggtgtcggt ggccaccaga 7440
gtctttctgg attccttcct gccaagatct gcaagatcaa cactgggatt gattgctaga 7500
gcagcagccc gagtttggaa cccatcaata cattttctgt ggtacaagct aggtgttttg 7560
agctaagagt taccaactaa gacagaggtt catcggaaag gaaacgggag taaaagaaag 7620
ggaggaggga gggaggggaa aagagagatg ggggaaggaa gagagacagg gaaggagaga 7680
gcagggtttc atttctgtcc ttctgtttcc aacttctgtt tggaaatgct gtttacttgg 7740
ggcgtcttgc ccgggatctt gggccaggga agtgccggcc tgaagtgacc ccctcttcct 7800
gtacttctct ccccgctctg ggccgcctcc gctcccscct cccccgcaca ggttgtacgg 7860
gatcaaatgc gccaagtgca gcatsggctt cagcaagaac gacttcgtga tgcgtgcccg 7920
ctccaaggtg tatcacatcg agtgtttccg ctgtgtggcc tgcagccgcc agctcatccc 7980
tggggacgaa tttgcgcttc gggaggacgg tctcttctgc cgagcagacc acgatgtggt 8040
ggagagggcc agtctaggcg ctggcgaccc gctcagtccc ctgcatccag cgcggccact 8100
gcaaatggca ggtactcctc tgcccggctc gggtaggcag gcgccaggtt aagccagcct 8160
gtgtgccagc ggccacaaca acyatggtag ctacaggggt ggtcgtagtg tttgcctgca 8220
gttaaatgaa gtgttctgta tgcaatttgc gctgtgctct gctcctttgc agcaaggttc 8280
aatgcactca ctgtctccct tgattccccg agcacaccta caccgtctgt gtgtctctat 8340
atggttacac ataaatgtac accacttgtg tacacgtgta tacacacgcc caaacattac 8400
ttccagttcg ctctggcctc caaaccttgg cttgctgaaa acgggcttca gctcccagcc 8460
aggtattctc ctgctgccta attaaagggg cggagccccg ggtccctgga gcttcatcct 8520
ttaacccaat gaaggaagct taggtggcct gaagtcattt agtctcccaa atccttttc 8580

cttgtgagtt gcttcacact cgaaattttt tttttaattt ttttatcttt ctgtgagaga 8640
acaggactga aaagatacag ttttaaaaac tgcaggccat tgcacagagt tgtaatataa 8700
aactgtcaac aagcttatct gcagtaattg ccttttaaag ggagcctgct tctttaaatc 8760
attcattcta tatgatttgg tgagaatttc atcttcaggc ccatggttgt agctctaaat 8820
tgaccccata ggtgttggcc tgaccctagg gggttgtaga aggtgcagga tttgtatcat 8880
gtagataaga ggactcattc ccaaggaaga ggagtggaaa cacagcaagg ttggccggga 8940
ccaaagcagt gggttagaag gtggacagtg tttccaaacc tgacttcctg ccatgaatag 9000
atctacccct ttgcagtttt aaagtatcaa ttcccactaa acactgaagg tgaggaaact 9060
atagccctcc cttacccttc tgccttctgg cagctctaag aattctgttc agggggattt 9120
gtgactagtt tgcaccgggg cacggctggg gtggtgctcc tgttcagtgg agcctgcact 9180
ctgcttgtgg ggaagcacag aggaagctaa aataccgaga gggaggcggg ggacatctcc 9240
cagccaccgt ttatctagag cctaggcagc tcaacagagt ttccgttttc cactgcttgg 9300
gatcagccca tctcaggaac atccatgtat taccttagat ttaatactaa gagcagggat 9360
tggagatatg gcagaaatag cgaatctctt cagccccttc acatgactgt cctctcggac 9420
tgaagttcaa ggcgttctgg cagagttctc gaccttcccc ttgcagaagt ccctgctggt 9480
gtagtattta tggctgtcac tgaagtgctc tgcgttcctt tccctggtac cctctgtggc 9540
cttggcccaa gagaaaattc tgatcctgga gagggtggta atcaatgtaa ctggggccca 9600
gtctgggcac aaggaaaggt gagaatggag gagaaacagt gctgaaaaat gccacccctg 9660
ctgtgaacag ggggacagac tttgagacct gcttcccttg gctaacactt tgttgacacg 9720
aggaggggcg agtgctgcgt ttcaggccgg gattactcag caaagacctc tgcagattag 9780
agaggaagat tttattctcc ctttcaccct cttcgccccc acctctgccg ccccctgctt 9840
tgtgtgctga ggctgcaaac cctagccatt gtcctgagta tctcgggcgg gcgagcaagt 9900
aagcgggcgg gcgggcgggc aagcgagcga gcgagcgagc gcgcgaccgc gggcgggccg 9960
gcaagcgagc ctccagccca gcgctcacgg cgctccttgc cccgcagcgg agcccatctc 10020
cgccaggcag ccagccctgc ggccccacgt ccacaagcag ccggagaaga ccacccgcgt 10080
gcggactgtg ctgaacgaga agcagctgca caccttgcgg acctgctacg ccgcaaaccc 10140
gcggccagat gcgctcatga aggagcaact ggtagagatg acgggcctca gtycccgtgt 10200
gatccgggtc tggtttcaaa acaagcggtg caaggacaag aagcgaagca tcatgatgaa 10260
gcaactccag cagcagcagc ccaatgacaa aactgtgagt ggctctgggg ccgggcaggg 10320
aatgcgaggg ggaaggagac gcagcgtgcg aggtgcgttc ctggtacgca ggatcgcacg 10380
gttttcaatc ctgctcctgg gcaggagttt ggccggggct gcccctcatc cttacccccc 10440
tacccatgcc ccgggggaca ggctacccgg cgccggccgc cagctgaggg cggggaagct 10500
gggaggctcc gtgcgccggg ggagcagcat ccaggtccca acctcgtggg tgggctcatg 10560
cccttccacc tcgcctgtac ctgtgaaccg gagaaacgcc gtcctcccct ctgagggcag 10620
gcggcaacga ggtttggccc gggttttgcc aacattcaga tcgtcagttc ctcacgtaca 10680
caagaagagg gagggataat accttggatt cctgcctaca tccaggggtt ccgtgggcag 10740
gtcaccctgt gagcccccag ggcgcaccgc acttctaagt aaggtcggcc gctgcgcctt 10800
caggctggcg agttccccca aggtgacccg catgcccaga tcaccctctg ctccaggtga 10860
agcccaggcc tccacagagg catcaggccc ctcgcaccag tatccactgt tatcttggtc 10920
ccacggaagc acccactctg caggcctcct ggtgaagtta agctagagtt tcttttcttc 10980
ctttttttct tttctttctt tttctttttt ttttttttt ttttttttt ttttttttt 11040
ttactgcttt ggacctattt ttaaatgcca taaaatctgc tgtcattaaa cttggcaggc 11100
tggccaagat tgggccaggg cactttctga gttggttagt gcataatagc acaataggaa 11160
ccagacccaa atgctttggg gggatggagt gggggggctgg ctcttccttg aggagaacgg 11220
cttggaaaaa atctgcagct aactgaaact gctcagaaaa ccaccctgtc tagaggctga 11280
agggaagccc tgcttacctc agctttttag ttctgggaag ctatggtctg agaaggcaga 11340
ggggaggaat tgggctgagc tgtgaaggta agggggaaga agaaaatcaa agtagaattt 11400
ggtttaataa ggtccatgca gacctaatag tccagcccac agaggcagaa aaacaaaaca 11460

ataaaacaaa ttgaattcta actaatatcc gtaggtacgg crgattaact gagtcaataa 11520
agaccactat atagataaga taataccagg gtatatttgc ttagcctgtg cagacaacgg 11580
agggagggaa tttgctcatt aacatgttgg gattggttgg ggggcctatt cacagaatat 11640
ccaggggatg acaggaactc ccatggtggc tgccagtcca gagagacacg acggtggctt 11700
acaggctaac ccagtggaag tacaaagtta ccagccacct tggaaagtac tgagcgactt 11760
cgccttgcag agtgacatag atcagcctgc ttttcagcaa ctggtaagtg tcagctccca 11820
gatggaagag gctgaattcc caacaggaga ctctggttta actgtcacac attgaaagat 11880
tcagtgggga gggtgccttc ttgggctcag ggttggggag aaaccaagga ggtgggtaat 11940
gaagagaagg gagacaaatg cagggaaaac gaacctcttg gcatcttttt ttttttaatg 12000
agactgcata atttgaccat ataggttgaa ttttctatca atcaggcctt ctttgaagga 12060
ttaatttcaa ggtacctaac tctaggtagc atgtgccaga agatgtacag tgttggagaa 12120
tcatacatct tagaatttta gagttgtcaa ggacttcagg aaatcgtctt ggcatttcaa 12180
tcagcaatta gtaagtttat ccttcctgag catctagaga atgggatata taggacccaa 12240
atcaaggcga ttgtagtata taatcaatgc tataatacac aatcagtgtt gttatcaata 12300
aacagcaggc atgtgtctgg gtacaatttt caaatatatt aataaagatt attcaaatag 12360
atgaatactt tttgttacag tatcccttgc tgggaatgtc ttaatctaaa atgtaggacc 12420
gtttaaatgt tttcaagtgt atgagttcaa atgtcataga gacacacagt gtgtaccatg 12480
tatagcaaaa gggacatgag ctctaccaat cagaagtaaa gtgtttaatc tgtgaaaacc 12540
ttaacatgtt ttccacatcc agagaggaga aaattaattc actttttgcc tacaaaaggc 12600
ttaaggggtc aagataaata agaacaataa atatatgtcc tttgtaatat gctatattta 12660
tatagatgat tttttttttct taaagagtaa tcagccttat agaatcttgt tttataaaat 12720
gtaaagatct atcctgaaac cttgttccct tttttggaaa tgaagcttta gttgaggtta 12780
gctttttacc ctcatattta cctggagggc atttgctttc tcaatgtcaa cagttaggta 12840
attggccaga ggcaagtggt taaaagggct tggccccagg cttgtgtttg caaatgctaa 12900
gtgggtgcag aggctagaag tcccttaatc tcatattgga aaaatttact gtagaaagaa 12960
atgtaggctc tagaactagg aaaaaaaaat tattctaagc tcattaatct gttgagttat 13020
ttgagcgaat cctgaatcac aggaggaagg taaggggagg cttcagggca gccaaatgtt 13080
tgcactttct gaaactttag tgtcagatga gagcagtgga agggaagctg aggcaggagt 13140
gggcatagtt agagaaggtt tacaacagca gtacaatgcg tttagggtta aaagaaggag 13200
tcagatattt aagaaggagt cagatattag ggttaaaaga aggagtcaga atgggatgat 13260
gtcataatat atgggtctca ttttggaagg aagagcctga tttaaagaga gagagagaga 13320
aaggccaaga ggcagcagga ccaacaagga agaatgccca agctgtgagc ctgctgagga 13380
gttaatcttt gttctgtgga gcctcctctc aatctcctgt caaaggatct gagcctgtta 13440
cggattttcc aactgaagaa gagagtcttt gatgcctaga gactgagagc tcacctactc 13500
ccagggcaac atgtagccag caggataatt ttatttcgag catgcatagt agagttgtga 13560
tgccatttta cagtgggaaa cacatttgtt cttaaataat ttaatgcaac ataatgttgg 13620
gaattcagtt tcagttaaaa cagagatctt ttggaagatg ggaaagtgag aggatttctt 13680
cccaagtttt tctcctctag gctttctcta agcctgttaa aattcagtta tctatgtgaa 13740
tatctttaca tatctatcta cacaaacatt tctacatata caatatgatg agtttataat 13800
ctttttatga atactattcc agtgtccttt atttrtttct caaccttcta tgcaggtcaa 13860
tttttcagaa ggaggaccgg gctctaattc cactggcagt gaagtagcat caatgtcctc 13920
tcaacttcca gatacaccta acagcatggt agccagtcct attgaggcat gaggaacatt 13980
cattctgtat ttttttttccc tgttggagaa agtgggaaat tataatgtcg aactctgaaa 14040
caaaagtatt taacraccca gtcaatgaaa actgaatcaa gaaatgaatg ctccatgaaa 14100
tgcacgaagt ctgttttaat gacaaggtga tatggtagca acactgtgaa gacaatcatg 14160
ggattttact agaattaaac aacaaacaaa acgcaaaacc cagtatatgc tattcaatga 14220
tcttagaagt actgaaaaaa aaagacgttt ttaaaacgta gaggatttat attcaaggat 14280
ctcaaagaaa gcattttcat ttcactgcac atctagagaa aaacaaaaat agaaaatttt 14340

ctagtccatc ctaatctgaa tggtgctgtt tctatattgg tcattgcctt gccaaacagg 14400
agctccagca aaagcgcagg aagagagact ggcctccttg gctgaaagag tcctttcagg 14460
aaggtggagc tgcattggtt tgatatgttt aaagttgact ttaacaaggg gttaattgaa 14520
atcctgggtc tcttggcctg tcctgtagct ggtttatttt ttactttgcc ccctccccac 14580
ttttttgag atccatcctt tatcaagaag tctgaagcga ctataaaggt ttttgaattc 14640
agatttaaaa accaacttat aaagcattgc aacaaggtta cctctatttt gccacaagcg 14700
tctcgggatt gtgtttgact tgtgtctgtc caagaacttt tcccccaaag atgtgtatag 14760
ttattggtta aaatgactgt tttctctctc tatggaaata aaaaggaaaa aaaaaaagga 14820
aacttttttt gtttgctctt gcattgcaaa aattataaag taatttatta tttattgtcg 14880
gaagacttgc cacttttcat gtcatttgac attttttgtt tgctgaagtg aaaaaaaaag 14940
ataaaggttg tacggtggtc tttgaattat atgtctaatt ctatgtgttt tgtctttttc 15000
ttaaatatta tgtgaaatca aagcgccata tgtagaatta tatcttcagg actatttcac 15060
taataaacat ttggcataga taaataaata aacacagtga tccagggtcc ctcttttatg 15120
tttacaaaca agctattcaa atgtctgaaa gtaaggattg catgtgaatt ggggtaaact 15180
gacacacagt aggggtgata ttgtattgca agtagttaag gtcgagtttg catgtttggg 15240
tttcagtagt tggatcaaag ttggttggaa gccccttagc tctcagctat gataatgcta 15300
ctggttttgt gaaaatgaga tgctgttctc tgtcctgccc ctctctttcc ttccccccagt 15360
ctctttcttg ccatacccca cccatccacc caccccaaat aggaagaggg tcttgcttaa 15420
aatggctttt gattcagagt cactatttta ttttcttaga atacgtgatt tataccaaag 15480
tctcctgctt gggattgttg aaatgaccca tatgtcaaat atataagctc ttggtaagag 15540
agactttctg cctagccagt gaggtatgac cattatttac cagtggcggc tgtcagccga 15600
ttgagaaatg ctgctctatt tgtgcagaag gaacaggcct agagaagaga gtatgctctg 15660
caattcccag actagtcggt gtatgcattt tcttgcctct tctcccctct cccacctccg 15720
cttcgtccct tgcccaagtc caatttcaaa aatcttggga agcatgactg gctatggaga 15780
gaaacatggc atatatactg ttgggaaacc acctttatg ggatggggta gcaggtcatg 15840
tgctttgctg gacccatcag ctatcagggt tacgggcatg ctctctggga cagactattc 15900
ctcccctggc atttagcaat agttcaagtg caagagccag aaccctggaa aagataaatg 15960
gcactgcccg aaatcgggct tcttggtgct tcttcatttg ctggcattga attaggattt 16020
ggggctccac cctgttcccc ctcctccctt taccccaccc caggatgatg ctttgactct 16080
ggaaactgca ttgatttgta tgcgcgagtg taatgccatt ttcagtacct gacaattctt 16140
attatgaaag ctaaagtacc ctaattgtgt gtcatgtggc gtacccctcc aggctcgata 16200
gggtgaggtg gaggctgtgc cccccagctc tcaagtcact ttcaccctct gcaaacctgc 16260
ctcaagtctg gaagccagcc tcaattaggc ttttgtttgg cttggttcta cataacgggt 16320
taattacagg ctctaacgtg ggggcagaga gaaaacaaac tttgcattga aaatgttaca 16380
gtgcattaga catgcccgtt gattactttg attgttgtgt ctaattcttg acagaggggt 16440
aataatttgg ctggtggcaa taatggcttg cagtcttctg ctgagaaaaa cagaagttat 16500
gtcttatctc acacggatct tctccgggaa aggtttttt gttgttgttt gtttgttttg 16560
ttttgttttg ttttcactc ggatctccct cctctaatct cttgccaagg gactttactc 16620
ttttagcttc ccattcacta tttctattgg tctcttttt ccttttcct caccatcata 16680
gctgaaatgg caactaatat gcatgtgtaa ggatggcagg ttcaacatga atattggtgt 16740
catgcctgca aagttactga gagtccaagc cttccttact tggaaaacac tggacccact 16800
tccccagccg tagcctctgg tggactcaaa aatgttaggg actagagatg gaaaacactt 16860
tatttaggat ttgaatgcaa tagagagact taagtgaggc acccatcttt gagaatcttt 16920
ttcctcctcg aattgatgaa ttgatttcta agattaagaa tatgtttgca tgtttgggtg 16980
gggggcgttg atgttttaca tacatagatg tcagtattct gaagtccatc gcttgaattt 17040
aagcacaata atttagggag ggggtgatga atgatttggt ttcacttgct agattcaaga 17100
tatttgttgc ctgatggctg agaatggatt gatgggtcaa tagcattcta gttcttttat 17160
tatttatacg aaggtgtgtc ttccccttt agactcctca gcctgcactc taaaatgtgt 17220

agctgaaatg tcattttgtt tgcaaatcat ctcttgcata cttttctctg cacacctctg 17280
gagactgagt tttgagctct gctcttgtaa tgaaattcct cttgttcgtt attatcttta 17340
atcgtccctc ccttattact ctgaattctt ccttctggac attatgactc atttcctcca 17400
gcttctgcaa gctaaaattt aatgcagcta aaaaacatct agtatgctcc ccttaaatca 17460
gaaaattcat atcaatatac atagtcttaa agtcaggata ttgtttcccc tgtagtttca 17520
gcgcaatgct atcttccatt tcagcctatt ttccctctat cccattcatc aaatgaactt 17580
gggttttctt acaaatcaaa aaagttcttt tctaagttag aagcatgttt gaagattttt 17640
ttctcctttg aaacaaggtt tctggttatt tgtaattctg tgtgtaccca cttgcccatt 17700
gacgacttgc ttggagttca cgattaaata gatgactttg tgatgcttgg ctgctattca 17760
ccctgatcaa cttgttaaac tcataatgtt gacctctttt ctggtgtatc ttatcctttt 17820
gagactagct aaaaagacaa tgtgctttca ggttacgttg gaagtacctt gtatttccct 17880
agcaccttga attttaagcc ataagatcta tctcatgaaa tgtaactgtt gtaaaatccc 17940
atgtggcttt acttagcgtt ttctttttte t 17971

<210> 2
<211> 1041
<212> DNA
<213> Homo sapiens

<400> 2
atgggagatc caccaaaaaa aaaacgtctg atttccctat gtgttggttg cggcaatcag 60
attcacgatc agtatattct gagggtttct ccggatttgg aatggcatgc ggcatgtttg 120
aaatgtgcgg agtgtaatca gtatttggac gagagctgta catgctttgt tagggatggg 180
aaaacctact gtaaaagaga ttatatcagg ttgtacggga tcaaatgcgc caagtgcagc 240
atcggcttca gcaagaacga cttcgtgatg cgtgcccgct ccaaggtgta tcacatcgag 300
tgtttccgct gtgtggcctg cagccgccag ctcatccctg gggacgaatt tgcgcttcgg 360
gaggacggtc tcttctgccg agcagaccac gatgtggtgg agagggccag tctaggcgct 420
ggcgacccgc tcagtcccct gcatccagcg cggccactgc aaatggcagc ggagcccatc 480
tccgccaggc agccagccct gcggccccac gtccacaagc agccggagaa gaccacccgc 540
gtgcggactg tgctgaacga gaagcagctg cacaccttgc ggacctgcta cgccgcaaac 600
ccgcggccag atgcgctcat gaaggagcaa ctggtagaga tgacgggcct cagtccccgt 660
gtgatccggg tctggtttca aaacaagcgg tgcaaggaca agaagcgaag catcatgatg 720
aagcaactcc agcagcagca gcccaatgac aaaactaata tccaggggat gacaggaact 780
cccatggtgg ctgccagtcc agagagacac gacggtggct tacaggctaa cccagtggaa 840
gtacaaagtt accagccacc ttggaaagta ctgagcgact tcgccttgca gagtgacata 900
gatcagcctg cttttcagca actggtcaat ttttcagaag gaggaccggg ctctaattcc 960
actggcagtg aagtagcatc aatgtcctct caacttccag atacacctaa cagcatggta 1020
gccagtccta ttgaggcatg a 1041

<210> 3
<211> 346
<212> PRT
<213> Homo sapiens

<400> 3
Met Gly Asp Pro Pro Lys Lys Lys Arg Leu Ile Ser Leu Cys Val Gly

```
1               5                  10                  15

Cys Gly Asn Gln Ile His Asp Gln Tyr Ile Leu Arg Val Ser Pro Asp
            20                  25                  30

Leu Glu Trp His Ala Ala Cys Leu Lys Cys Ala Glu Cys Asn Gln Tyr
        35                  40                  45

Leu Asp Glu Ser Cys Thr Cys Phe Val Arg Asp Gly Lys Thr Tyr Cys
    50                  55                  60

Lys Arg Asp Tyr Ile Arg Leu Tyr Gly Ile Lys Cys Ala Lys Cys Ser
65                  70                  75                  80

Ile Gly Phe Ser Lys Asn Asp Phe Val Met Arg Ala Arg Ser Lys Val
                85                  90                  95

Tyr His Ile Glu Cys Phe Arg Cys Val Ala Cys Ser Arg Gln Leu Ile
            100                 105                 110

Pro Gly Asp Glu Phe Ala Leu Arg Glu Asp Gly Leu Phe Cys Arg Ala
        115                 120                 125

Asp His Asp Val Val Glu Arg Ala Ser Leu Gly Ala Gly Asp Pro Leu
    130                 135                 140

Ser Pro Leu His Pro Ala Arg Pro Leu Gln Met Ala Ala Glu Pro Ile
145                 150                 155                 160

Ser Ala Arg Gln Pro Ala Leu Arg Pro His Val His Lys Gln Pro Glu
                165                 170                 175

Lys Thr Thr Arg Val Arg Thr Val Leu Asn Glu Lys Gln Leu His Thr
            180                 185                 190

Leu Arg Thr Cys Tyr Ala Ala Asn Pro Arg Pro Asp Ala Leu Met Lys
        195                 200                 205

Glu Gln Leu Val Glu Met Thr Gly Leu Ser Pro Arg Val Ile Arg Val
    210                 215                 220

Trp Phe Gln Asn Lys Arg Cys Lys Asp Lys Lys Arg Ser Ile Met Met
225                 230                 235                 240

Lys Gln Leu Gln Gln Gln Gln Pro Asn Asp Lys Thr Asn Ile Gln Gly
                245                 250                 255

Met Thr Gly Thr Pro Met Val Ala Ala Ser Pro Glu Arg His Asp Gly
```

260 265 270

Gly Leu Gln Ala Asn Pro Val Glu Val Gln Ser Tyr Gln Pro Pro Trp
275 280 285

Lys Val Leu Ser Asp Phe Ala Leu Gln Ser Asp Ile Asp Gln Pro Ala
290 295 300

Phe Gln Gln Leu Val Asn Phe Ser Glu Gly Gly Pro Gly Ser Asn Ser
305 310 315 320

Thr Gly Ser Glu Val Ala Ser Met Ser Ser Gln Leu Pro Asp Thr Pro
325 330 335

Asn Ser Met Val Ala Ser Pro Ile Glu Ala
340 345

<210> 4
<211> 15
<212> DNA
<213> Homo sapiens

<400> 4
cagcccgrgc ttcgg 15

<210> 5
<211> 15
<212> DNA
<213> Homo sapiens

<400> 5
ctgcgtcrga ccaat 15

<210> 6
<211> 15
<212> DNA
<213> Homo sapiens

<400> 6
taaggaasag aggtg 15

<210> 7
<211> 15
<212> DNA

<213> Homo sapiens

<400> 7
aagagagrtg cccga 15

<210> 8
<211> 15
<212> DNA
<213> Homo sapiens

<400> 8
ttcaccarct gtaca 15

<210> 9
<211> 15
<212> DNA
<213> Homo sapiens

<400> 9
tattttayct tgtgg 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapiens

<400> 10
accttgtrgg gctcg 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapiens

<400> 11
cgctcccscc tcccc 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapiens

<400> 12

gcagcatsgg cttca 15

<210> 13
<211> 15
<212> DNA
<213> Homo sapiens

<400> 13
caacaacyat ggtag 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapiens

<400> 14
cctcagtycc cgtgt 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapiens

<400> 15
gtacggcrga ttaac 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapiens

<400> 16
tttatttrtt tctca 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapiens

<400> 17
atttaacrac ccagt 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapiens

<400> 18
gcccggcagc ccgrg 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapiens

<400> 19
tctccccccga agcyc 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapiens

<400> 20
cgcgcgctgc gtcrg 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapiens

<400> 21
atcgccattg gtcyg 15

<210> 22
<211> 15
<212> DNA
<213> Homo sapiens

<400> 22
aagagataag gaasa 15

<210> 23
<211> 15
<212> DNA

<213> Homo sapiens

<400> 23
ctcgggcacc tctst 15

<210> 24
<211> 15
<212> DNA
<213> Homo sapiens

<400> 24
ataaggaaga gagrt 15

<210> 25
<211> 15
<212> DNA
<213> Homo sapiens

<400> 25
cgcggctcgg gcayc 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapiens

<400> 26
gggctgttca ccarc 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapiens

<400> 27
ggtggttgta cagyt 15

<210> 28
<211> 15
<212> DNA
<213> Homo sapiens

<400> 28

agaggctatt ttayc 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapiens

<400> 29
cgagccccac aagrt 15

<210> 30
<211> 15
<212> DNA
<213> Homo sapiens

<400> 30
tattttacct tgtrg 15

<210> 31
<211> 15
<212> DNA
<213> Homo sapiens

<400> 31
gcacaccgag cccya 15

<210> 32
<211> 15
<212> DNA
<213> Homo sapiens

<400> 32
cgcctccgct cccsc 15

<210> 33
<211> 15
<212> DNA
<213> Homo sapiens

<400> 33
tgtgcggggg aggsg 15

<210> 34
<211> 15
<212> DNA
<213> Homo sapiens

<400> 34
ccaagtgcag catsg 15

<210> 35
<211> 15
<212> DNA
<213> Homo sapiens

<400> 35
tcttgctgaa gccsa 15

<210> 36
<211> 15
<212> DNA
<213> Homo sapiens

<400> 36
cggccacaac aacya 15

<210> 37
<211> 15
<212> DNA
<213> Homo sapiens

<400> 37
ctgtagctac catrg 15

<210> 38
<211> 15
<212> DNA
<213> Homo sapiens

<400> 38
gacgggcctc agtyc 15

<210> 39
<211> 15
<212> DNA

<213> Homo sapiens

<400> 39
cggatcacac gggra 15

<210> 40
<211> 15
<212> DNA
<213> Homo sapiens

<400> 40
ccgtaggtac ggcrg 15

<210> 41
<211> 15
<212> DNA
<213> Homo sapiens

<400> 41
gactcagtta atcyg 15

<210> 42
<211> 15
<212> DNA
<213> Homo sapiens

<400> 42
gtgtccttta tttrt 15

<210> 43
<211> 15
<212> DNA
<213> Homo sapiens

<400> 43
gaaggttgag aaaya 15

<210> 44
<211> 15
<212> DNA
<213> Homo sapiens

<400> 44

aaaagtattt aacra 15

<210> 45
<211> 15
<212> DNA
<213> Homo sapiens

<400> 45
tcattgactg ggtyg 15

<210> 46
<211> 10
<212> DNA
<213> Homo sapiens

<400> 46
cggcagcccg 10

<210> 47
<211> 10
<212> DNA
<213> Homo sapiens

<400> 47
cccccgaagc 10

<210> 48
<211> 10
<212> DNA
<213> Homo sapiens

<400> 48
gcgctgcgtc 10

<210> 49
<211> 10
<212> DNA
<213> Homo sapiens

<400> 49
gccattggtc 10

<210> 50
<211> 10
<212> DNA
<213> Homo sapiens

<400> 50
agataaggaa 10

<210> 51
<211> 10
<212> DNA
<213> Homo sapiens

<400> 51
gggcacctct 10

<210> 52
<211> 10
<212> DNA
<213> Homo sapiens

<400> 52
aggaagagag 10

<210> 53
<211> 10
<212> DNA
<213> Homo sapiens

<400> 53
ggctcgggca 10

<210> 54
<211> 10
<212> DNA
<213> Homo sapiens

<400> 54
ctgttcacca 10

<210> 55
<211> 10
<212> DNA

<213> Homo sapiens

<400> 55
ggttgtacag 10

<210> 56
<211> 10
<212> DNA
<213> Homo sapiens

<400> 56
ggctatttta 10

<210> 57
<211> 10
<212> DNA
<213> Homo sapiens

<400> 57
gccccacaag 10

<210> 58
<211> 10
<212> DNA
<213> Homo sapiens

<400> 58
tttaccttgt 10

<210> 59
<211> 10
<212> DNA
<213> Homo sapiens

<400> 59
caccgagccc 10

<210> 60
<211> 10
<212> DNA
<213> Homo sapiens

<400> 60

ctccgctccc 10

<210> 61
<211> 10
<212> DNA
<213> Homo sapiens

<400> 61
gcgggggagg 10

<210> 62
<211> 10
<212> DNA
<213> Homo sapiens

<400> 62
agtgcagcat 10

<210> 63
<211> 10
<212> DNA
<213> Homo sapiens

<400> 63
tgctgaagcc 10

<210> 64
<211> 10
<212> DNA
<213> Homo sapiens

<400> 64
ccacaacaac 10

<210> 65
<211> 10
<212> DNA
<213> Homo sapiens

<400> 65
tagctaccat 10

<210> 66
<211> 10
<212> DNA
<213> Homo sapiens

<400> 66
gggcctcagt 10

<210> 67
<211> 10
<212> DNA
<213> Homo sapiens

<400> 67
atcacacggg 10

<210> 68
<211> 10
<212> DNA
<213> Homo sapiens

<400> 68
taggtacggc 10

<210> 69
<211> 10
<212> DNA
<213> Homo sapiens

<400> 69
tcagttaatc 10

<210> 70
<211> 10
<212> DNA
<213> Homo sapiens

<400> 70
tcctttattt 10

<210> 71
<211> 10
<212> DNA

<213> Homo sapiens

<400> 71
ggttgagaaa 10

<210> 72
<211> 10
<212> DNA
<213> Homo sapiens

<400> 72
agtatttaac 10

<210> 73
<211> 10
<212> DNA
<213> Homo sapiens

<400> 73
ttgactgggt 10

<210> 74
<211> 1620
<212> DNA
<213> Homo sapiens

<220>
<221> allele
<222> (30)
<223> PS1: polymorphic base A or G

<220>
<221> misc_feature
<222> (61)..(120)
<223> N's represent sequence between PS1 and PS2

<220>
<221> allele
<222> (150)
<223> PS2: polymorphic base A or G

<220>
<221> misc_feature
<222> (181)..(240)
<223> N's represent sequence between PS2 and PS3

<220>
<221> allele
<222> (270)
<223> PS3: polymorphic base G or C

<220>
<221> misc_feature
<222> (301)..(360)
<223> N's represent sequence between PS3 and PS4

<220>
<221> allele
<222> (390)
<223> PS4: polymorphic base G or A

<220>
<221> misc_feature
<222> (421)..(480)
<223> N's represent sequence between PS4 and PS5

<220>
<221> allele
<222> (510)
<223> PS5: polymorphic base A or G

<220>
<221> misc_feature
<222> (541)..(600)
<223> N's represent sequence between PS5 and PS6

<220>
<221> allele
<222> (630)
<223> PS6: polymorphic base C or T

<220>
<221> misc_feature
<222> (661)..(720)
<223> N's represent sequence between PS6 and PS7

<220>
<221> allele
<222> (750)
<223> PS7: polymorphic base G or A

<220>
<221> misc_feature

<222> (781)..(840)
<223> N's represent sequence between PS7 and PS8

<220>
<221> allele
<222> (870)
<223> PS8: polymorphic base C or G

<220>
<221> misc_feature
<222> (901)..(960)
<223> N's represent sequence between PS8 and PS9

<220>
<221> allele
<222> (990)
<223> PS9: polymorphic base C or G

<220>
<221> misc_feature
<222> (1021)..(1080)
<223> N's represent sequence between PS9 and PS10

<220>
<221> allele
<222> (1110)
<223> PS10: polymorphic base T or C

<220>
<221> misc_feature
<222> (1141)..(1200)
<223> N's represent sequence between PS10 and PS11

<220>
<221> allele
<222> (1230)
<223> PS11: polymorphic base C or T

<220>
<221> misc_feature
<222> (1261)..(1320)
<223> N's represent sequence between PS11 and PS12

<220>
<221> allele
<222> (1350)
<223> PS12: polymorphic base G or A

<220>
<221> misc_feature
<222> (1381)..(1441)
<223> N's represent sequence between PS12 and PS13

<220>
<221> allele
<222> (1470)
<223> PS13: polymorphic base A or G

<220>
<221> misc_feature
<222> (1501)..(1560)
<223> N's represent sequence between PS13 and PS14

<220>
<221> allele
<222> (1590)
<223> PS14: polymorphic base G or A

<400> 74
ggccagagcc agaacagccc ggcagcccgr gcttcggggg agaacggcct gagccccgag 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
gctcgcctga cgtccccgcg cgctgcgtcr gaccaatggc gatggagctg agttggagca 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
gcagagaagt ttgagtaaga gataaggaas agaggtgccc gagccgcgcc gagtctgccg 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
gaagtttgag taagagataa ggaagagagr tgcccgagcc gcgccgagtc tgccgccgcc 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
ccagcatcct ctctgtgggc tgttcaccar ctgtacaacc accatttcac tgtggacatt 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
gatccaccaa aaagtaagag gctattttay cttgtggggc tcggtgtgct gttcttgtgc 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
ccaaaaagta agaggctatt ttaccttgtr gggctcggtg tgctgttctt gtgcggggtt 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
tctccccgct ctgggccgcc tccgctcccs cctcccccgc acaggttgta cgggatcaaa 900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
tacgggatca aatgcgccaa gtgcagcats ggcttcagca agaacgactt cgtgatgcgt 1020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1080
ccagcctgtg tgccagcggc cacaacaacy atggtagcta caggggtggt cgtagtgttt 1140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
agcaactggt agagatgacg ggcctcagty cccgtgtgat ccgggtctgg tttcaaaaca 1260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320
gaattctaac taatatccgt aggtacggcr gattaactga gtcaataaag accactatat 1380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1440
tatgaatact attccagtgt cctttatttr tttctcaacc ttctatgcag gtcaattttt 1500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1560
tgtcgaactc tgaaacaaaa gtatttaacr acccagtcaa tgaaaactga atcaagaaat 1620

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 February 2002 (14.02.2002)

PCT

(10) International Publication Number
WO 02/012498 A3

(51) International Patent Classification[7]: C12N 15/12, 5/10, C12Q 1/68, C07K 14/475, 16/22, G06F 17/00

(21) International Application Number: PCT/US01/24664

(22) International Filing Date: 6 August 2001 (06.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KLIEM, Stefanie, E.** [DE/DE]; Kiefernweg 37, D-61440 Oberusel (DE). **KOSHY, Beena** [IN/US]; 1298 Hartford Turnpike, Apt. 11B, North Haven, CT 06473 (US). **TANGUAY, Debra, A.** [US/US]; 902 Aspen Glen Drive, Hamden, CT 06518 (US).

(74) Agent: **SHANER, Sandra, L.**; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
28 August 2003

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 02/012498 A3

(54) Title: HAPLOTYPES OF THE ISL1 GENE

(57) Abstract: Novel genetic variants of the ISL1 Transcription Factor, LIM/Homeodomain, (islet-1) (ISL1) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the ISL1 gene. Compositions and methods for haplotyping and/or genotyping the ISL1 gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/24664

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C12N5/10 C12Q1/68 C07K14/475 C07K16/22
G06F17/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, PAJ, WPI Data, SEQUENCE SEARCH, BIOSIS, MEDLINE, CHEM ABS Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | RIGGS A C ET AL: "Characterization of the LIM/homeodomain gene islet-1 and single nucleotide screening in NIDDM." DIABETES. UNITED STATES JUN 1995, vol. 44, no. 6, June 1995 (1995-06), pages 689-694, XP002226953 ISSN: 0012-1797 the whole document --- | 1-24,32 |
| Y | DATABASE [Online] 17 September 1999 (1999-09-17) DOE JOINT GENOME INSTITUTE STANFORD HUMAN GENOME CENTER: "Homo sapiens chromosome 5 clone CTD-2314G24, complete sequence" retrieved from EM_HUM Database accession no. AC010478 XP002226954 abstract --- -/-- | 1-24,32 |

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 26 February 2003 | 26. 06. 2003 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax: (+31-70) 340-3016 | Hagenmaier, S |

Form PCT/ISA/210 (second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/24664

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | US 5 972 614 A (BENTLEY KEVIN L ET AL)<br>26 October 1999 (1999-10-26)<br>the whole document | 1-24,32 |
| Y | RUANO: "Genaissance awarded broad patent on methods for collecting genes for variation analysis across populations"<br>GENAISSANCE PRESS RELEASE, [Online]<br>8 November 1999 (1999-11-08), XP002229209<br>Retrieved from the Internet:<br><URL:WWW.biomedicine.ch/publications/genaissance/g111199.html><br>[retrieved on 2003-02-26]<br>abstract | 1-24,32 |
| Y | JUDSON R ET AL: "THE PREDICTIVE POWER OF HAPLOTYPES IN CLINICAL RESPONSE"<br>PHARMACOGENOMICS, ASHLEY PUBLICATIONS, GB,<br>vol. 1, no. 1, February 2000 (2000-02),<br>pages 15-26, XP001113104<br>ISSN: 1462-2416<br>the whole document | 1-24,32 |
| Y | CLARK A G ET AL: "HAPLOTYPE STRUCTURE AND POPULATION GENETIC INFERENCES FROM NUCLEOTIDE- SEQUENCE VARIATION IN HUMAN LIPOPROTEIN LIPASE"<br>AMERICAN JOURNAL OF HUMAN GENETICS, UNIVERSITY OF CHICAGO PRESS, CHICAGO,, US,<br>vol. 63, 1998, pages 595-612, XP002937239<br>ISSN: 0002-9297<br>cited in the application<br>the whole document | 1-24,32 |
| Y | ULBRECHT M ET AL: "Assoziation of beta 2-adrenoreceptor variants with bronchial hyperresponsiveness"<br>AMERICAN JOURNAL OF RESPIRATORY AND CRITICAL CARE MEDICINE, AMERICAN LUNG ASSOCIATION, NEW YORK, NY, US,<br>vol. 161, no. 2, February 2000 (2000-02),<br>pages 469-474, XP002188594<br>ISSN: 1073-449X<br>cited in the application<br>the whole document | 1-24,32 |

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | RUANO G ET AL: "GENOTYPING AND HAPLOTYPING OF POLYMORPHISMS DIRECTLY FROM GENOMIC DNA VIA COUPLED AMPLIFICATION AND SEQUENCING (CAS)" NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB, vol. 19, no. 24, 1991, pages 6877-6882, XP000197814 ISSN: 0305-1048 the whole document --- | 1-24,32 |
| A | CLAIBORNE STEPHENS J: "SINGLE-NUCLEOTIDE POLYMORPHISMS, HAPLOTYPES AND THEIR RELEVANCE TO PHARMACOGENTICS" MOLECULAR DIAGNOSIS, NAPERVILLE, IL, US, vol. 4, no. 4, 1999, pages 309-317, XP002937520 ISSN: 1084-8592 the whole document --- | |
| A | WO 93 05783 A (GEN HOSPITAL CORP) 1 April 1993 (1993-04-01) the whole document --- | |
| A | MAREZ D ET AL: "Polymorphism of the cytochrome P450 CYP2D6 gene in a European population: characterization of 48 mutations and 53 alleles, their frequencies and evolution" PHARMACOGENETICS, CHAPMAN & HALL, LONDON, GB, vol. 7, no. 3, 1997, pages 193-202, XP002955538 ISSN: 0960-314X the whole document --- | |
| A | EMILIEN G ET AL: "IMPACT OF GENOMICS ON DRUG DISCOVERY AND CLINICAL MEDICINE" QUARTERLY JOURNAL OF MEDICINE, CLARENDON PRESS, OXFORD, GB, vol. 93, no. 7, June 2000 (2000-06), pages 391-423, XP000941207 ISSN: 0033-5622 the whole document --- | |
| A | LU ANTONY Y H: "Drug-metabolism research challenges in the new millennium: Individual variability in drug therapy and drug safety." DRUG METABOLISM AND DISPOSITION, vol. 26, no. 12, December 1998 (1998-12), pages 1217-1222, XP002223866 ISSN: 0090-9556 the whole document --- | |

-/--

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/24664

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| P,Y | DRYSDALE C M ET AL: "COMPLEX PROMOTER AND CODING REGION BETA2-ADRENERGIC RECEPTOR HAPLOTYPES ALTER RECEPTOR EXPRESSION AND PREDICT IN VIVO RESPONSIVENESS" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 97, no. 19, 12 September 2000 (2000-09-12), pages 10483-10488, XP002940094 ISSN: 0027-8424 the whole document --- | 1-24,32 |
| T | WO 01 80156 A (WINDEMUTH ANDREAS ;GENAISSANCE PHARMACEUTICALS IN (US); STEPHENS J) 25 October 2001 (2001-10-25) the whole document ----- | |

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/24664

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 31
because they relate to subject matter not required to be searched by this Authority, namely:

Claim 31 refers to subject matter which the ISA is not required to search (R 39.1 (vi) PCT).

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-24. 32 (all partially)

**Remark on Protest**

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises a guanine at PS1; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of a guanine or an adenine at PS1; method for the identification of an association between a trait and an ISL1 sequence comprising guanine at PS1; a genome anthology comprising such an ISL1 sequence.

2. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises a guanine at PS2; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of a guanine or an adenine at PS2; method for the identification of an association between a trait and an ISL1 sequence comprising guanine at PS2; a genome anthology comprising such an ISL1 sequence.

3. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises a cytosine at PS3; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of a cytosine or a guanine at PS3; method for the identification of an association between a trait and an ISL1 sequence comprising cytosine at PS3; a genome anthology comprising such an ISL1 sequence.

4. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises an adenine at PS4; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of an adenine or a guanine at PS4; method for the identification of an association between a trait and an ISL1 sequence comprising adenine at PS4; a genome anthology comprising such an ISL1 sequence.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

5. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises a guanine at PS5; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of a guanine or an adenine at PS5; method for the identification of an association between a trait and an ISL1 sequence comprising guanine at PS5; a genome anthology comprising such an ISL1 sequence.

6. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises a thymine at PS6; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of a thymine or a cytosine at PS6; method for the identification of an association between a trait and an ISL1 sequence comprising thymine at PS6; a genome anthology comprising such an ISL1 sequence.

7. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises an adenine at PS7; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of an adenine or a guanine at PS7; method for the identification of an association between a trait and an ISL1 sequence comprising adenine at PS7; a genome anthology comprising such an ISL1 sequence.

8. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises a guanine at PS8; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of a guanine or a cytosine at PS8; method for the identification of an association between a trait and an ISL1 sequence comprising cytosine at PS8; a genome anthology comprising such an ISL1 sequence.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

9. Claims: 1-30,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises a guanine at PS9; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of a guanine or a cytosine at PS9; method for the identification of an association between a trait and an ISL1 sequence comprising guanine at PS9; a genome anthology comprising such an ISL1 sequence; an isolated ISL1 polynucleotide comprising a fragment of Seq.ID 2, wherein the fragment comprises a guanine at PS9; a recombinant non-human organism transformed or transfected with said polynucleotide; an isolated polypeptide comprising methionine at a position corresponding to amino acid position 81 of Seq.ID 3; an isolated monoclonal antibody specific for said polypeptide; a method for screening for drugs targeting said polypeptide.

10. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises a cytosine at PS10; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of a cytosine or a thymine at PS10; method for the identification of an association between a trait and an ISL1 sequence comprising cytosine at PS10; a genome anthology comprising such an ISL1 sequence.

11. Claims: 1-30,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises a thymine at PS11; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of a thymine or a cytosine at PS11; method for the identification of an association between a trait and an ISL1 sequence comprising thymine at PS11; a genome anthology comprising such an ISL1 sequence; an isolated ISL1 polynucleotide comprising a fragment of Seq.ID 2, wherein the fragment comprises a thymine at PS11; a recombinant non-human organism transformed or transfected with said serine at a position corresponding to amino acid position 219 of Seq.ID 3; an isolated monoclonal antibody specific for said polypeptide; a method for screening for drugs targeting said polypeptide.

12. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises an adenine at PS12; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of an adenine or a guanine at PS12; method for the identification of an association between a trait and an ISL1 sequence comprising adenine at PS12; a genome anthology comprising such an ISL1 sequence.

13. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises a guanine at PS13; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of a guanine or an adenine at PS13; method for the identification of an association between a trait and an ISL1 sequence comprising guanine at PS13; a genome anthology comprising such an ISL1 sequence.

14. Claims: 1-24,32 (all partially)

An isolated ISL1 polynucleotide comprising a fragment of Seq.ID 1, wherein the fragment comprises an adenine at PS14; a recombinant non-human organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the ISL1 locus comprising the detection of an adenine or a guanine at PS14; method for the identification of an association between a trait and an ISL1 sequence comprising adenine at PS14; a genome anthology comprising such an ISL1 sequence.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 01/24664

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| US 5972614 A | 26-10-1999 | US 5866404 A | 02-02-1999 |
| | | US 6221588 B1 | 24-04-2001 |
| | | US 2002132348 A1 | 19-09-2002 |
| WO 9305783 A | 01-04-1993 | AU 2492192 A | 27-04-1993 |
| | | CA 2116130 A1 | 01-04-1993 |
| | | EP 0666745 A1 | 16-08-1995 |
| | | JP 7501518 T | 16-02-1995 |
| | | WO 9305783 A1 | 01-04-1993 |
| WO 0180156 A | 25-10-2001 | AU 5372001 A | 30-10-2001 |
| | | EP 1290613 A1 | 12-03-2003 |
| | | WO 0180156 A1 | 25-10-2001 |